

# Strukturmodell des Terminationsmoduls der Surfactin Synthetase

## *Crystal Structure of the Surfactin Synthetase Termination Module*



**Der Hauptteil dieser Arbeit wurde im folgenden Artikel publiziert:**

Tanovic, A., Samel, S. A., Essen, L. O., and Marahiel, M. A. **(2008)**, Crystal structure of the termination module of a nonribosomal peptide synthetase, *Science*, **321**: 659-663.

**Im Verlauf der Promotion wurde weiterhin folgender Artikel veröffentlicht:**

Strieker, M., Tanovic, A., and Marahiel, M. A. **(2010)** Nonribosomal peptide synthetases: structures and dynamics, *Curr Opin Struct Biol*, **20**: 234-240.

# STRUKTURMODELL DES TERMINATIONSMODULS DER SURFACTIN SYNTHETASE

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich Chemie der Philipps-Universität Marburg (Hochschulkennziffer: 1180)

vorgelegt von

*Dipl. Chem. Alan Tanović*

geb. in Georgsmarienhütte

Erstgutachter: Professor Dr. Mohamed A. Marahiel

Zweitgutachter: Professor Dr. Klaus Reuter

Prüfungsbeisitzer: Professor Dr. Ulrich Tallarek

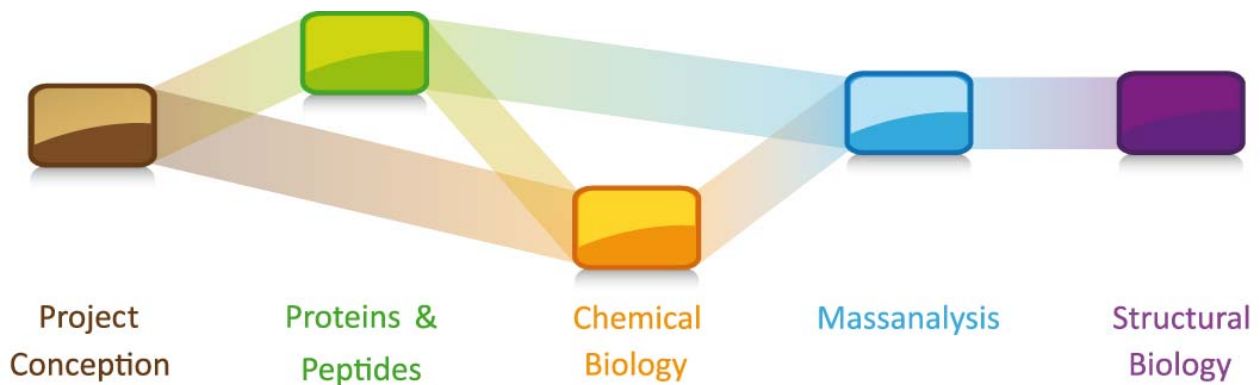
Vom Fachbereich Chemie der Philipps-Universität Marburg als Dissertation

am: 22.03.2012 angenommen.

Erfolgreich verteidigt am: 27.04.2012

Marburg, 2012.

# LEBENS LAUF



## Lebenslauf

### Promotionsstudium

07/2007 – 04/2012	Fachbereich Chemie, Abteilung für Biochemie der Philipps-Universität Marburg
Arbeitsgruppe	Prof. Mohamed A. Marahiel
Titel	<i>Strukturmodell des Terminationsmoduls der Surfactin Synthetase</i>
Beschreibung	Untersuchungen der Architektur multimodularer nichtribosomaler Peptid-synthetasen, der für die Produktsynthese essentiellen Interaktionen beteiligter Domänen sowie der dadurch induzierten Änderungen der globalen Struktur

### Studium

06/2007	Diplomarbeit am Fachbereich Chemie, Abteilung für Biochemie der Philipps-Universität Marburg
Arbeitsgruppe	Prof. Mohamed A. Marahiel
Titel	<i>Rationale Mutagenese nichtribosomaler Peptidsynthetasen und erste strukturelle Untersuchungen konformationserstarrter Module</i>
Beschreibung	Kristallographische untersuchung an unimodularen konformations-erstarrten nichtribosomalen Peptidsynthetasen

### Publikationen

Forschungsbericht	<b>Tanovic A</b> , Samel SA, Essen LO, Marahiel MA (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase <i>Science</i> <b>321</b> (5889): 659-63
Übersichtsartikel	Strieker M, <b>Tanović A</b> , Marahiel MA (2010). Nonribosomal peptide synthetases: structures and dynamics <i>Curr Opin Struct Biol</i> <b>20</b> (2): 234-40



# CURRICULUM VITAE

## Contact Data:

Name: Alan Tanović  
Email: alan.tanovic@gmx.de  
Skype: alan.tanovic  
Xing: [http://www.xing.com/profile/Alan\\_Tanovic](http://www.xing.com/profile/Alan_Tanovic)



## Curriculum Vitae >

### Doctoral Studies

07/2007 – 04/2012      Doctoral Studies at the Biochemistry Department of the Faculty of Chemistry, Philipps-University of Marburg

Working group      Prof. Mohamed A. Marahiel

Doctoral Thesis      *Crystal Structure of the Surfactin Synthetase Termination Module*

Description      Investigation of the architecture of multimodular NRPS and of the interactions of involved and for the product synthesis essential domains and thereby induced conformational switches in the global structure.

### Higher Education

06/2007      University studies in chemistry, specialization in biochemistry at the Philipps-University Marburg

Working group      Prof. Mohamed A. Marahiel

Diploma Thesis      *Rational mutagenesis and first structural studies of conformational frozen modules of the Nonribosomal peptide synthetases*

Description      Study of unimodular conformational frozen nonribosomal peptide synthetases as potential crystallization targets

### Publications

Research Article      **Tanovic A**, Samel SA, Essen LO, Marahiel MA (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase  
*Science* **321**(5889): 659-63

Review      Strieker M, **Tanović A**, Marahiel MA (2010). Nonribosomal peptide synthetases: structures and dynamics  
*Curr Opin Struct Biol* **20**(2): 234-40

# ZUSAMMENFASSUNG

---

Nichtribosomale Peptidsynthetasen (NRPS) zeigen, trotz der hohen Anzahl ihrer Bausteine und der somit großen Vielfalt ihrer Produkte, eine konstante strukturelle Verwandtschaft. Die NRPS-Systeme sind modular aufgebaut, wobei definierte Domänen die einzelnen Reaktionen katalysieren. Neben den katalytischen Domänen ist die Rolle des dynamischen Peptidyl-Carrier-Proteins (PCP) für das Verständnis der NRPS-Wirkungsweise entscheidend. Studien vergleichbarer Systeme zeigen, dass die das Carrier-Protein umgebenden Linkerregionen aus Prolin- und Alanin-reichen Sequenzen bestehen, wodurch die Bildung von Sekundärstrukturen erschwert und die Dynamik dieser Domänen unterstützt wird.

Erst die in dieser Arbeit durchgeführte Arretierung der dynamischen PCP-Domäne durch rationale Mutagenese ermöglichte die Beobachtung des Kristallwachstums und somit die erste Strukturaufklärung eines vollständigen NRPS-Moduls. Die Struktur des Terminationsmoduls der Surfactin Synthetase (SrfAC-S1003A) zeigt die Interaktion der dynamischen PCP-Domäne mit der katalytischen Kondensationsdomäne (C-Domäne). Das SrfAC-Strukturmodell beschreibt die erste Struktur einer  $^D C_L$ -Domäne. Desweiteren konnte die Bildung einer engen Interaktion zwischen der Kondensations- und den zwei Subdomänen der Adenylierungsdomäne ( $A_{core}$ - und  $A_{sub}$ -Domäne), die als *Workbench* bezeichnet wird, festgestellt werden. Dabei nimmt die  $A_{sub}$ -Domäne eine bisher nicht beobachtete Konformation ein. Die C- und  $A_{core}$ -Domäne repräsentieren 83% der Masse eines Elongationsmoduls (C-, A- und PCP-Domäne), auf deren starrem Gerüst die dynamische  $A_{sub}$ - und PCP-Domäne interagieren. Anschließende Untersuchungen mittels der Kleinwinkel-Streuung (SAXS) zeigten, dass diese enge Interaktion auch im solvatisierten Zustand bestehen bleibt. Somit konnte die wichtige Rolle der *Workbench*, auch als Modell-System für das generelle Verständniss der NRPS-Systeme, bestätigt werden.

Die Strukturaufklärung des vollständigen NRPS-Moduls ermöglicht die eindeutige Definition der flankierenden PCP-Linkerbereiche. Ihre sequenzielle Analyse zeigt, dass diese Bereiche aus Prolin- und Alanin-reichen Sequenzen bestehen und somit nicht zur Bildung von Sekundärstrukturen neigen. Dagegen teilt der  $\alpha$ -helikale Linkerbereich zwischen der C- und A-Domäne nicht die Affinität zu Prolin- und Alanin-reichen Sequenzen und ist zur Bildung von Sekundärstrukturen fähig.

Zusätzlich konnte in dem SrfAC-S1003A-Strukturmodell die spezifische Interaktion zweier Synthetasen über ein Helix-Hand-Motiv gezeigt werden. Dabei interagiert die C-terminale  $\alpha$ -Helix (COM-Helix) der ersten Synthetase mit dem in die C-Domäne integrierten Hand-Motiv der nachfolgenden Synthetase. Dadurch ist die spezifische Erkennung der Synthetasen untereinander gewährleistet, wodurch die Synthesereihenfolge erhalten bleibt (*assembly line*).

Durch die Kombination mit der Bidomänen-Struktur aus TycC konnte desweiteren ein multimodulares NRPS-Modell erstellt werden. Dieses Modell zeigt eine linksläufige Schraubachse, bei der jedes Modul entlang der *Assembly-Line* um  $120^\circ$  gedreht ist. Weitere Untersuchungen mittels CytoEM an trimodularen NRPS-Systemen sollen diese Modul-Anordnung bestätigen. Neben der globalen Architektur der NRPS, zeigt das Akzeptor-Modell auch die ersten strukturell aufgeklärten Linkerbereiche einer *swinging domain*. Die in dieser Arbeit erstmals aufgeklärte Struktur eines NRPS-Moduls ermöglicht neue synthetische und chemoenzymatische Ansätze für rationales Design und *protein engineering* an den nichtribosomalen Peptidsynthetasen und so den Zugang zu neuen biologisch wirksamen Peptidverbindungen.

## A B S T R A C T

---

Nonribosomal peptide synthetases (NRPSs) utilize complex regiospecific and stereospecific reactions to assemble structurally and functionally diverse peptides. Despite the large number of their components and the great diversity of products, the NRPS systems hold a close structural relationship. The NRPS systems shows a modular organisation with defined domains which catalyzing the single reactions. In addition to the catalytic domains the role of dynamic peptidyl carrier protein (PCP) is crucial for understanding of the NRPS mode of action.

In this work the disruption of the PCP dynamic via serine to alanin variation or attached pantotheine derivatives and whose effect on the global structure was investigated. Freezing the PCP domain in the AH state allowed the structure determination of a complete termination module of the surfactin synthetase. This structure model shows various unique properties, describing the domain interactions between the condensation domain and PCP, the first  $^D C_L$  domain structure model, an artificial COM helix – COM hand motif as well as a novel orientation of the adenylation subdomain relating to the adenylation coredomain. Furthermore, the formation of a close interaction between the condensation and the two subdomains of the adenylation domain ( $A_{core}$  and  $A_{sub}$ ), which is called the workbench be noted. Subsequent studies using small-angle scattering (SAXS) showed that this close interaction persists even in the solvated state. This studies confirm the important role of the workbench as a model system for the general understanding of NRPS systems.

By combining the structural data of the bidomain structure of TycC<sub>5-6</sub> with the surfactin termination module SrfAC a multimodulare NRPS model can be created. This model shows a counter clockwise screw axis where each module along the assembly line ist rotated by 120° as well the complete structural model of the different linker regions.

The knowledge about the NRPS organization increase the possibility of constructing active chimeric enzymes, where intact modules or even single domains are exchanged to allow the biosynthesis of novel nonribosomal peptides and facilitate the access to new biologically active peptide compounds.



## Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Liponsäure und Biotin.....	2
1.1.1	Der Pyruvat-Dehydrogenase-Komplex (PDH).....	3
1.1.2	Pyruvat-Carboxylase Komplex (PC).....	5
1.2	Coenzym A.....	7
1.2.1	Fettsäuresynthase (FAS).....	8
i.)	Architektur eukaryontischer Fettsäuresynthasen in Säugetieren.....	10
ii.)	Architektur eukaryontischer Fettsäuresynthasen aus Pilzen.....	13
1.2.2	Polyketidsynthase (PKS).....	15
i.)	Architektur der Polyketidsynthasen.....	16
1.2.3	Nichtribosomale Peptidsynthetase (NRPS).....	18
i.)	Die Adenylierungsdomäne (A-Domäne).....	22
ii.)	Peptidyl Carrier Protein (PCP).....	25
iii.)	Die Kondensationsdomäne (C-Domäne).....	28
iv.)	Die Thioesterase (TE-Domäne).....	30
v.)	Modifizierende 'in cis'-Domänen (Substitution).....	33
vi.)	Modifizierende 'in cis'-Domänen (Insertion).....	35
vii.)	Modifizierende 'in trans'-Domänen.....	37
2	Aufgabenstellung.....	39
3	Material.....	41
3.1	Equipment.....	41
3.2	Chemikalien, Enzyme und Verbrauchsmaterial.....	42
3.3	Verwendete Software und Datenbanken.....	44
3.4	Verwendete Oligonukleotid-Primer.....	45

3.5 Vektor-Systeme.....	46
3.6 Mikroorganismen.....	47
3.7 Nährmedien.....	47
3.7.1 LB-Medium.....	47
3.7.2 M9-Medium.....	47
3.7.3 TB-Medium.....	48
4 Methoden.....	49
4.1 Molekularbiologische Methoden.....	49
4.2 Proteinchemische Methoden.....	49
4.2.1 Genexpression im präparativen Maßstab.....	50
4.2.2 Zellaufschluß.....	51
4.2.3 Proteinaufreinigung.....	51
4.2.4 Gelfiltrations-Chromatographie.....	52
4.2.5 HI-TRAP Umpufferungsverfahren.....	53
4.2.6 Ankonzentrieren von Proteinlösungen.....	53
4.2.7 SDS-Polyacrylamidgelelektrophorese (SDS-PAGE).....	54
4.2.8 Proteinkonzentrationsbestimmung.....	55
i.) Qualitative Proteindetektion nach Bradford.....	55
ii.) Spektrometrische Proteinkonzentrationsbestimmung.....	55
iii.) Proteinkonzentrationsbestimmung mittels Refraktometer.....	55
4.3 Biochemische Methoden.....	56
4.3.1 Adenylierungsaktivitätsmessungen durch den ATP/PPi-Test.....	56
4.3.2 PCP-Beladung mittels Panthoteinierungsreaktion.....	56
4.3.3 Allosterische Inhibitionsassays.....	57
4.3.4 Gekoppelte Amino-CoA und Aminosäure Beladung.....	57

# INHALTSVERZEICHNIS

---

4.4 Bioanalytische Methoden.....	58
4.4.1 MALDI-TOF.....	58
4.4.2 Hybrid-Massenspektrometrie (HPLC- und LC-MS).....	58
4.4.3 Bottom-Up Proteinanalyse.....	59
4.5 Chemische Synthesen.....	60
4.5.1 Peptidfestphasensynthese.....	60
4.5.2 Amino-CoA Synthese.....	61
4.6 Methoden zur Aufklärung makromolekularer Strukturen.....	64
4.6.1 Dynamische Lichtstreuung unterstützte Kleinwinkelstreuung.....	64
i.) Datensammlung.....	67
ii.) Datenreduktion und Analyse.....	67
iii.) Modelaufbau und Verifikation.....	67
4.6.2 Einzelpartikel-Cryo-Elektronenmikroskopie.....	68
i.) Probenvorbereitung und Datensammlung.....	69
ii.) Klassifizierung und Modellaufbau .....	70
4.6.3 Kristallographie.....	70
i.) Datensammlung und -prozessierung.....	72
ii.) Strukturlösung und Verfeinerung.....	73
ii.a)MAD – multiwavelength anomalous diffraction.....	74
ii.b)MIR – multiple/single isomorphous replacement.....	74
ii.c)MIRAS – Kombination aus isomorphen Ersatz und anomaler Dispersion .....	75
ii.d)MR – molecular replacement.....	75
4.7 Bioinformatische Methoden.....	75
5 Ergebnisse.....	77
5.1 Kristallisation unimodularer NRPS.....	77
5.1.1 Heterologe Genexpression von SrfAC-S1003A.....	78

5.1.2 Kristallisation und Datensammlung von SrfAC S1003A.....	79
5.1.3 Strukturlösung und Verfeinerung von SrfAC S1003A.....	81
5.2 Das SrfAC-S1003A-Strukturmodell.....	83
5.2.1 Katalytische und dynamische Domänen.....	85
i.) Kondensationsdomäne.....	88
ii.) Adenylierungsdomäne.....	90
iii.) Peptidyl-Carrier-Protein (PCP) und die Thioesterase.....	92
5.2.2 Intramodulare Wechselwirkungen.....	93
5.2.3 Intermodulare Wechselwirkung.....	98
5.3 Modell multimodularer NRPS-Systeme.....	102
5.4 Probenpräparation und Analyse.....	105
5.5 SAXS mit modularen NRPS.....	109
5.5.1 SrfAC – Workbench Untersuchung mittels SAXS .....	109
5.5.2 Kleinwinkelstreuung an dimodularen NRPS (TycB <sub>2-3</sub> ).....	111
6 Diskussion.....	116
6.1 NRPS Linker- und Intercore-Bereiche mit SrfAC als Templat.....	117
6.2 SrfAC Strukturmodell als NRPS-Strukturtemplat.....	121
6.2.1 PCP-Domäne: Arretierung durch rationale Mutagenese.....	122
6.2.2 Orientierung der ASub-Domäne als Indikator der Adenylierungsreaktion.....	124
6.2.3 Kondensationsdomäne als NRPS-Schnittstelle.....	126
i.) Oberflächenrepräsentationen und Kavitäten der Kondensationsdomäne.....	129
ii.) Analogien zwischen den Acetyltransferasen und den NRPS Kondensationsdomänen.....	132
6.3 Interaktion in multimodularen Systemen.....	137



# INHALTSVERZEICHNIS

---

6.4 Überblick: Die Disziplinen der NRPS.....	139
7 Ausblick.....	142
7.1 Multimodulare NRP-Synthetasen aus thermophilen Organismen.....	142
7.2 Cryo-EM an multimodularen NRPS.....	143
8 Anhang.....	148
8.1 Quellcode des Python-Scripts.....	148
8.2 Eingabedateien.....	151
8.2.1 Sequenz-Datei (Alignment).....	151
8.2.2 Control-Datei.....	152
8.3 Ausgabedatei.....	152
8.4 Alignment-Analyse.....	153
9 Abkürzungsverzeichnis.....	156
10 Literatur.....	158



# 1 Einleitung

Das Konzept der *swinging domains* und *swinging arms*[1] ermöglicht in multifunktionalen Enzymen und Enzym-Komplexen durch das Prinzip der Substrattranslokation die Synthese einzigartiger Intermediate und Produkte. Während der Substrattranslokation ist das Intermediat kovalent an der prosthetischen Gruppe (*swinging arm*) gebunden und ermöglicht so in multifunktionalen Systemen die Produktsynthese durch *active-site coupling* und *substrate channeling*. Der prosthetische Arm ist dabei selbst kovalent über die Seitenkette der konservierten Aminosäure an die dynamische Domäne (*swinging domain*) gebunden. Durch diese Kombination kann die spezifische Synthesereaktion durch die Wechselwirkung der dynamischen Domäne mit ihren Interaktionspartnern an verschiedenen aktiven Zentren und in einer bestimmten Reihenfolge erfolgen (*active-site coupling*). Die Synthesestrategie erfolgt stets am gebundenen Substrat und ermöglicht so die gezielte Synthese eines Produktes (*substrate channeling*), ohne unerwünschte Nebenreaktionen. Die prominentesten Vorläufer der *swinging arms* sind Liponsäure[2], Biotin[3] und Coenzym A[4] (Abb. 1), die durch ihre korrespondierenden *swinging domains* Lipoyl-Domäne (LD), Biotin-Carboxyl-Carrier-Protein (BCCP) und die Carrier-Proteine (Acyl-CP, Aryl-CP und Peptidyl-CP; ACP, ArCP und PCP) in vielen multifunktionalen Systemen, unter anderem der Pyruvat-Dehydrogenase (PDH), Pyruvat-Carboxylase (PC) und der Fettsäure- (FAS) und Polyketid-Synthase (PKS) sowie der in dieser Arbeit untersuchten nichtribosomalen Peptidsynthetasen (NRPS) eine essentielle Rolle spielen.

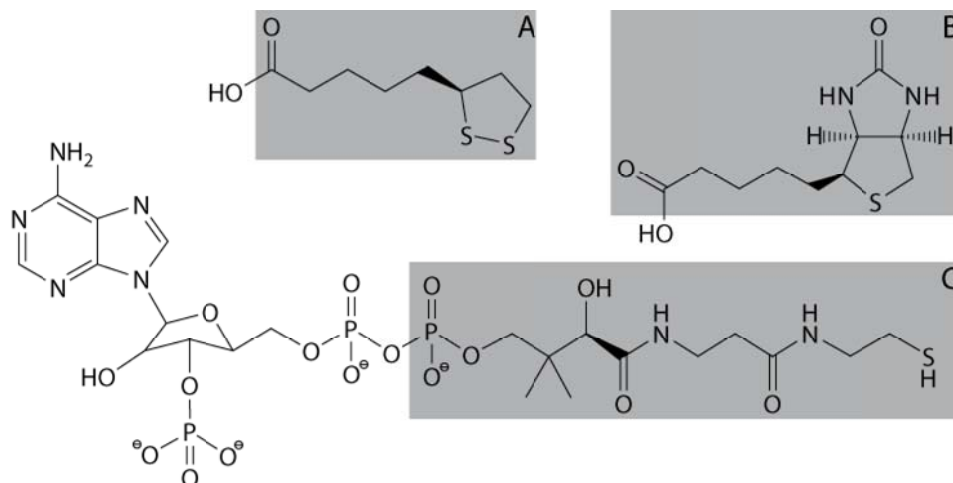


Abbildung 1: Vorläufer prosthetischer Gruppen: (A) Liponsäure, (B) Biotin und (C) Coenzym A.

## 1.1 Liponsäure und Biotin

Die Liponsäure spielt eine wichtige Rolle im primären Metabolismus, daher haben viele Organismen mehrere Biosynthesewege etabliert, um diese prosthetische Gruppe in die entsprechende Komplexe einzubauen. In der Zelle selbst liegt das Liponat nur in sehr geringer Konzentration als freie Säure vor, da das meiste Liponat auf die  $\epsilon$ -Aminogruppe eines konservierten Lysinrestes der Lipoyl-akzeptierenden Domäne des Zielkomplexes übertragen wird[5]. Die Insertion der Thiolgruppen an der C<sub>6</sub>- und C<sub>8</sub>-Position wird durch das Enzym LipA katalysiert, das durch genomische und biochemische Studien als Mitglied der Radikal-SAM Superfamilie identifiziert wurde[6]. Die Rolle und Funktion der Lipoyl-abhängigen Domänen sowie ihrer Interaktionspartner wurde besonders für die Dehydrogenase-Komplexe ausführlich beschrieben[7]. Diese Multienzym-Komplexe zählen zu der 2-Oxosäure-Dehydrogenase-Komplex-Familie, sind an vielen Reaktionen des oxidativen Metabolismus beteiligt und bestehen aus multiplen Kopien drei verschiedener Enzyme: die Thiamindiphosphat-abhängige (ThDP) Decarboxylase (E1), die Liponat-abhängige Acyltransferase (E2) und die Flavin-abhängige Dihydrolipoyl-Dehydrogenase (E3)[8]. Der vollständige Komplex kann eine Größe von über 9 MDa erreichen, während die assoziierten Untereinheiten unkonventionelle enzymatische Aktivitäten erlauben. Drei verwandte  $\beta$ -Hydroxy-Carbonsäure-Dehydrogenase-Komplexe zeigen Affinität für Pyruvat (PDH)[9],  $\beta$ -Hydroxyglutarat (OGDH)[10-11] und verzweigte  $\beta$ -Hydroxy-Carbonsäuren (BCDH)[12]. Mehrere der Liponsäure-abhängigen Systeme sind bereits strukturell aufgeklärt. Der Pyruvat-Dehydrogenase-Komplex (PDH) wurde mittels Kleinwinkelstreuung (SAXS)[13], Cryo-Elektronenmikroskopie (Cryo-EM)[14] und kristallographischer Methoden, als Vertreter der Lipoyl-Domäne-abhängigen Komplexe untersucht. Die Strukturmodelle einzelnen Domänen des Glycin-Decarboxylase-Komplexes (GCS)[15-16] wurden als Vertreter der H-Protein-abhängigen Komplexe mittels kristallographischer Methoden aufgeklärt. Die dynamischen Domänen (Lipoyl-Domäne und H-Protein) haben an der  $\epsilon$ -Aminogruppe des konservierten Lysins den Lipoamid-Arm als prosthetische Gruppe gebunden, der die Intermediate zwischen den einzelnen aktiven Zentren transportiert.

Biotin ist ein bizyklischer Naturstoff, bestehend aus einem Tetrahydroimidazolidon und einem Tetrahydrothiophen Ring, der mit der Seitenkette der Valeratsäure verknüpft ist

(Abb. 1)[17]. Im letzten Schritt der Biosynthese wird die Schwefelinsertion durch ein Enzym (BioB) der Radikal-SAM-Superfamilie katalysiert[3]. Das Enzym BioB zeigt signifikante Sequenzsimilarität zu der Liponat-Synthase (LipA) was durch genomische und biochemische Studien bestätigt wurde. Das Biotin wird durch die Biotin-Protein-Ligase (BPL) über eine Amidbindung zwischen der Valerat-Seitenkette und der  $\epsilon$ -Aminogruppe eines konservierten Lysinrestes kovalent an das Enzym gebunden. Diese posttranslationale Modifikation ist außerordentlich spezifisch für die Mitglieder aus der Gruppe der "Biotin-abhängigen Familie"[18]. Die Familie der Biotin-abhängigen Enzyme lässt sich in drei Hauptgruppen, jeweils nach ihrer Aktivität, unterteilen: die Gruppe der Carboxylasen[19], der Decarboxylasen[20] und der Transcarboxylasen[21]. Alle Enzyme die diesen Gruppen angehören sind an verschiedenen Stoffwechselwegen beteiligt.

## 1.1.1 Der Pyruvat-Dehydrogenase-Komplex (PDH)

Der Lipoyl-abhängige Pyruvat-Dehydrogenase-Komplex spielt eine zentrale Rolle im oxidativen Metabolismus der Zelle. Das zentrale Gerüst dieses Komplexes bildet die sogenannte E2-Einheit (E2p für die Pyruvat-, E2o für die  $\beta$ -Hydroxyglutarat- und E2b für die verzweigte  $\beta$ -Hydroxy-Carbonsäurespezifität), die in drei strukturell und funktional unabhängige Domänen unterteilt ist[1]. Der C-terminale Teil, mit einer CAT-Faltung (Chloramphenikol-Acetyltransferase) und einer Masse von 28 kDa, katalysiert den Transfer eines Acyl-Restes auf den ACP gebundenen Phosphopantothien-Arm und bildet ein Homotrimer aus (Abb. 2A) dessen Interaktionsflächen die Acyltransferase-Aktivität ausbilden[22]. Je nach Organismus bilden die als Homotrimer organisierten Einheiten ein oktaedrisches (24-mer) bzw. ikosaedrisches (60-mer) Innengerüst (E2 core). Jede Acyltransferase-Domäne des Innengerüsts enthält zwei exponierte Domänen: die 80 Aminosäuren lange, N-terminale, Lipoyl-akzeptierende Domäne (LD)[23] und die sehr kleine, nur aus ca. 35 Aminosäuren bestehende, periphere Untereinheit Bindedomäne (PSBD)[24], die durch die Interaktion zu den E1- und E3-Domänen[25-26] den vollständigen Dehydrogenase-Komplex ausbildet. Da jede E2-Domäne eine PSB-Domäne beinhalten kann, wird ein ikosaedrischer Dehydrogenase-Komplex von bis zu 60 Peripher-Enzymen und E1- und E3-Domänen, umgeben. Die Lipoyl-akzeptierende Domäne (PD) wird posttranslational an der  $\epsilon$ -Aminogruppe des konservierten

## STRUKTURMODELL DES TERMINATIONSMODULS DER SURFACTIN SYNTHETASE

Lysinrestes als Lipoylamid beladen. Die Alanin- und Prolin-reiche Linkerregion der PSB-Domäne ermöglicht der dynamischen Lipoyldomäne (LD) zwischen den einzelnen aktiven Zentren der drei partizipierenden Domänen (E1, E2 und E3) zu agieren (Abb. 2B). Dabei wirkt die prosthetische Lipoylgruppe als Träger der kovalent gebundenen Acyl-Intermediate und reduzierender Äquivalente (Abb. 2B).

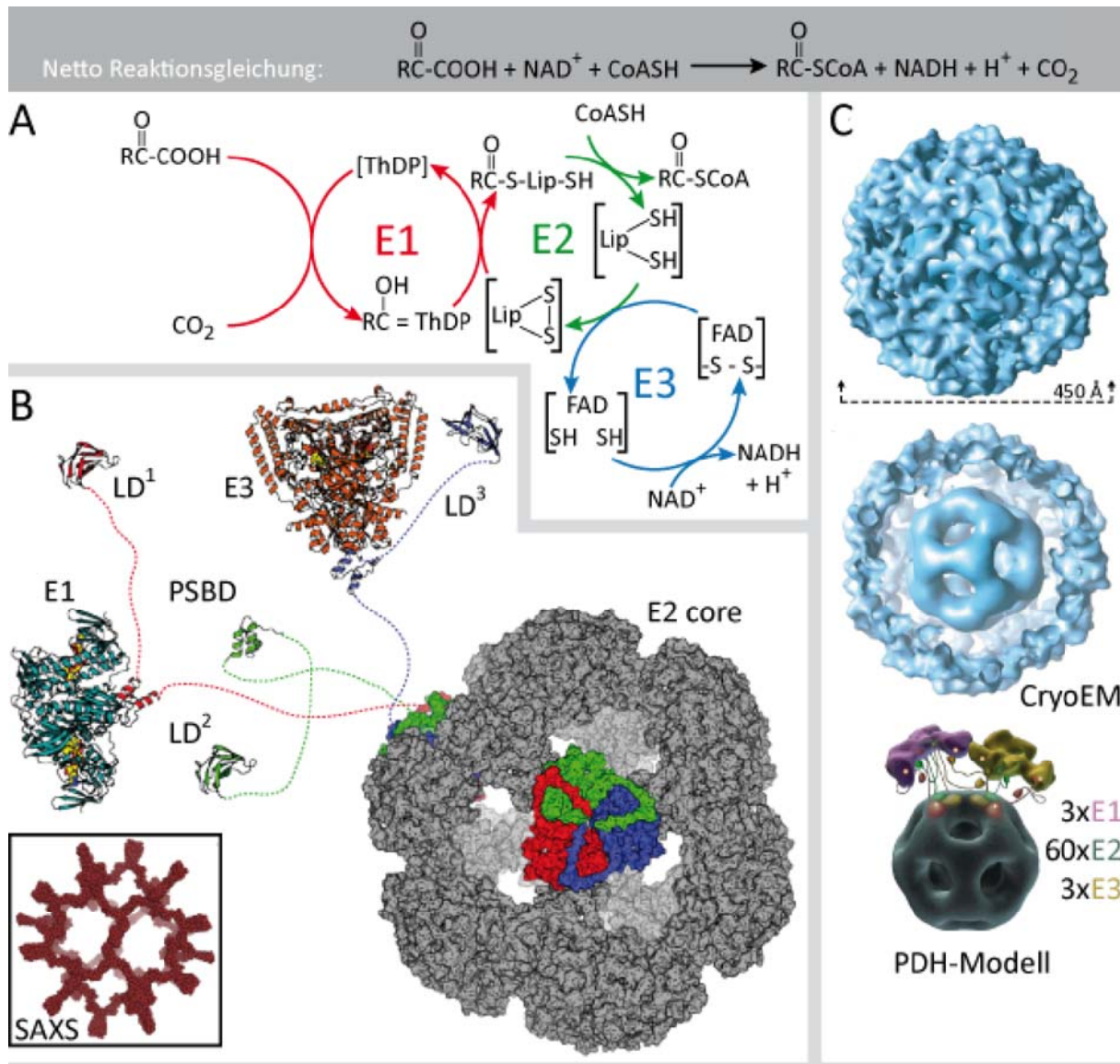


Abbildung 2: Strukturmodell und Reaktionsschema des Pyruvat-Dehydrogenase-Komplexes. (A) Reaktionen des PDH-Komplexes. (B) Strukturmodell des PDH-Komplexes aus *B. stearothermophilus* bestehend aus den Kristallstrukturen der Untereinheiten E1 (pdb: 1W85), E2 (pdb: 1B5S) und E3 (pdb: 1EBD), wobei E1 und E3 jeweils im Komplex mit der „peripheral subunit-binding domain“ (PSBD<sub>solo</sub>-pdb: 1W3D) kokristallisiert wurden, und der NMR-Struktur der Lipoamid-Arm-tragenden Lipoyl-Domäne (LD-pdb: 1LAB) sowie die Modell-Bestätigung durch SAXS (humane PDH)[13]. (C) Oberflächendarstellung des gesamten Komplexes und des inneren E2-Kerns nach der durch CryoEM ermittelten Elektronendichte[14]. Das nach den Strukturdaten erstellte PDH-Modell zeigt die aktiven Zentren sowie die katalytischen Domänen als Trimere.

Die PSB-Domäne ist die kleinste bekannte, zur Disulfidbrücken- bzw. Metallionenfreien Faltung fähige, Domäne. Diese Domäne ist für die Komplexierung der ThDP-abhängigen E1-Untereinheit sowie, mit Ausnahme der eukaryotischen PDH-Komplexe, für die Verbindung zu dem Dehydrogenase Dimer (E3) verantwortlich. In eukaryotischen PDH-Komplexen ist die PSBD an einer zusätzlichen E2-Untereinheit, dem sogenannten X-Protein, oder E3 Bindeprotein (E3BP), lokalisiert (Abb. 2B – SAXS)[13]. Die PSBD besteht aus zwei parallelen  $\alpha$ -Helices sowie einem Loop und bildet eine sehr starke Interaktion mit der E1-Domäne bzw. mit dem E3-Dimer aus. Die vier katalytischen Schritte des Pyruvat-Dehydrogenase-Komplexes sind in Abbildung 2B dargestellt. Das Substrat (Pyruvat) wird einer irreversiblen Decarboxylierung unterzogen, die durch die Thiamindiphosphat-abhängige Decarboxylase (E1) katalysiert wird. Hierbei entsteht ein reaktives ThDP-Enamin-Intermediat. Im zweiten Schritt findet eine reduktive Acylierung der Lipoylgruppe des beweglichen Lipoamid-Armes an dem  $\alpha_2\beta_2$ -E1-Heterotetramer statt. Die reduktive Acylierung erfolgt alternierend an zwei, 44 Å voneinander entfernten, aktiven Zentren. Die Lipoyl-Domäne wandert anschließend zu der Acetyltransferase (E2), wo die Acylgruppe auf die Thiolgruppe des Coenzym A übertragen wird. Im letzten Schritt der Katalyse interagiert die Lipoyl-Domäne mit dem aktiven Zentrum des Dehydrogenase-Dimers (E3), damit der Dithiolan-Ring des Lipoamid-Armes regeneriert und gleichzeitig  $\text{NAD}^+$  reduziert werden kann[27]. Cryo-EM-Untersuchungen am gesamten PDH-Komplexen zeigten, dass diese einen Durchmesser von ca. 450 Å, mit einem innerem Kern aus 60 E2-Einheiten aufweisen. Die dynamische Lipoyl-Domäne überwindet dabei den 75 - 90 Å langen Abstand zwischen dem inneren Kern und dem äußeren Ring um mit den aktiven Zentren der 42-48 E1-, 60 E2- oder 6-12 E3-Einheiten zu interagieren (Abb. 2C)[14].

## 1.1.2 Pyruvat-Carboxylase Komplex (PC)

Ein kristallographisch untersuchtes, Biotin-abhängiges Enzym ist die Pyruvat-Carboxylase (PC) aus *Rhizobium etli*[28]. Das PC-Tetramer besteht aus zwei Ebenen, die wiederum aus zwei aufeinanderliegenden, antiparallel orientierten, ca. 130 kDa schweren Polypeptidketten gebildet werden (Abb. 3A). Jede besteht aus drei funktionellen Domänen: die N-terminale Biotin-Carboxylase (BC), die zentrale Carboxyltransferase (CT) und das C-terminale Biotin-Carboxyl-Carrier-Protein (BCCP) sind in Abbildung 3 blau, gelb und rot dargestellt[29].

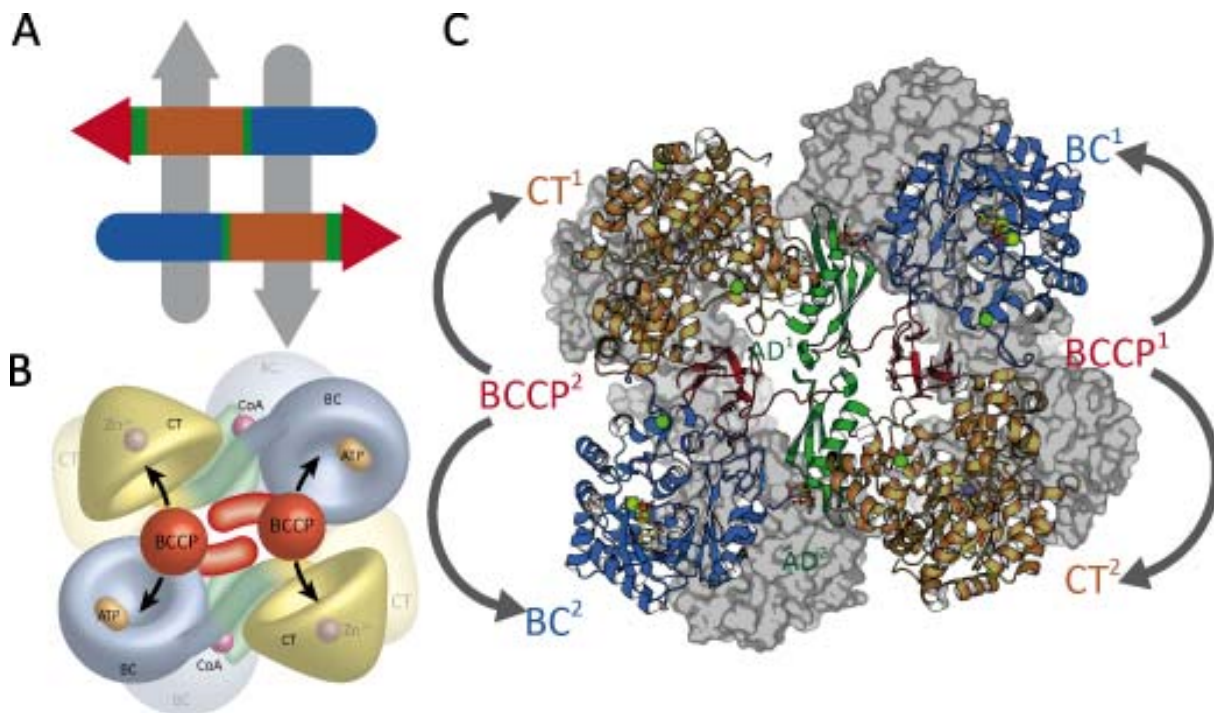


Abbildung 3: Schema, Modell und Strukturdarstellung des Pyruvat-Carboxylase-Komplexes. (A) Schema über die Anordnung der PC-Polypeptidketten. (B) Modell des Tetramers um die Bewegung der BCCP-Domäne zu der gegenüberliegenden Polypeptidkette zu verdeutlichen[28]. (C) Kristallstrukturdarstellung der Pyruvat-Carboxylase (pdb: 2QF7).

Das Enzym katalysiert mittels der kovalent gebundenen prosthetischen Gruppe Biotin die ATP abhängige Carboxylierung des Pyruvats zu Oxalacetat in zwei Stufen. Das am BCCP<sup>1</sup> gebundene Biotin wird zuerst am aktiven Zentrum der BC<sup>1</sup>-Domäne in einer ATP- und Bicarbonat-abhängigen Reaktion carboxyliert. Um die Carboxyl-Gruppe auf Pyruvat zu übertragen, muss das entstandene Carboxy-Biotin mit dem im aktiven Zentrum der CT<sup>2</sup>-Domäne der gegenüberliegenden Polypeptidkette liegenden Pyruvat interagieren. Die BCCP<sup>1</sup>-Domäne übernimmt den Transfer der reaktiven Intermediate zwischen den entsprechenden aktiven Zentren, wobei die Distanz intermolekular 65 Å und intramolekular 80 Å beträgt (Abb. 3B). Die allosterische Domänen (Abb. 3C, grün) können Acetyl-CoA binden und die Reaktion positiv beeinflussen. Sie unterliegen allosterischer Kontrolle. Diese erst im Strukturmodell entdeckte allosterische Domäne dient dabei als Vermittler zwischen den BC-, CT- und BCCP-Domänen und besteht aus vier antiparallelen  $\beta$ -Strängen. Das Pyruvat-Carboxylase-Strukturmodell besteht aus einem asymmetrischen Tetramer, einem Bidimer. In dem „oberen“ Teil, in der Abbildung 3C durch Sekundärstrukturelemente gezeigt (*cartoon*), findet eine Interaktion



zwischen der allosterischen Domäne (AD) und Acetyl-CoA statt. Dadurch ist eine räumliche Nähe zwischen den für die BCCP<sup>1</sup>-Domäne relevanten aktiven Zentren BC<sup>1</sup> und CT<sup>2</sup> zu beobachten. Im Gegenteil dazu kann im „unterem“ Teil des Bidimers, in der Abbildung 3C als Oberfläche gezeigt (*surface*) keine Interaktion mit Acetyl-CoA festgestellt werden, so daß die BCCP-Domänen delokalisiert vorliegen und daher keine konkrete Elektronendichte beobachtet werden konnte. Somit scheint die Bindung von Acetyl-CoA die BC-Domänen-Bewegung zu verursachen, die die beiden aktiven Zentren (BC<sup>1</sup> und CT<sup>2</sup> sowie BC<sup>2</sup> und CT<sup>1</sup>) in die Reichweite der Biotin-tragenden BCCP-Domäne bringt.

## 1.2 Coenzym A

Die als Coenzym A aktivierte Phosphopantethein-Arm wird mittels einer Mg<sup>2+</sup>-abhängiger Phosphopantotheinyl-Transferase (PPTase) kovalent an die Seitenkette des konservierten Serinrestes eines Trägerproteins (*carrier protein*) übertragen[30]. Die posttranslationale Phosphopantetheinylierung ermöglicht die Initiation, Elongation, Prozessierung sowie Abspaltung des kovalent gebundenen Acyl-, Aryl- bzw. Aminoacyl-S-Enzym-Thioesters. PPTasen werden sowohl nach ihren Sequenzhomologien als auch ihren Substratspezifitäten in drei Gruppen unterteilt[31]. Die erste Gruppe ist nach dem Enzym AcpS (*acyl carrier protein synthase*) aus *E. coli* benannt[32]. Die dieser Gruppe angehörigen Enzyme bestehen aus etwa 120 Aminosäuren und sind in fast allen Mikroorganismen für die Modifikation von ACPs der Fettsäuresynthasen (FAS) als auch der Typ II Polyketidsynthasen (PKS) verantwortlich[33]. Die 228 Aminosäuren lange PPTase Sfp aus *Bacillus subtilis* (Abb. 4) ist der Prototyp der zweiten Gruppe[34-35]. Die Enzyme dieser Gruppe wurden meistens assoziiert mit Genclustern der nichtribosomalen Peptidsynthetasen gefunden. Im Sfp-Strukturmodell[36] bildet die N-terminale Subdomäne eine ähnliche Faltung wie die AcpS-ähnliche C-terminale Subdomäne des Enzyms. Die beiden Subdomänen definieren so eine Faltung als Pseudodimer (*pseudo 2-fold dimer*; Abb. 4). Die Interaktionsfläche, an der das CoA bindet, wird bei Sfp durch die zwei Teile des Pseudodimeres und bei AcpS zwischen den einzelnen Monomeren des AcpS-Trimers gebildet. Diese wird von den in jeder Hälfte des Sfp Pseudodimeres enthaltenen, drei antiparallel angeordneten  $\beta$ -Strängen mit einer dazu diagonal angeordneten  $\alpha$ -Helix gebildet. Dabei werden der Purin-Rest und der benachbarte

Diphosphat-Rest des CoA-Moleküls durch mehrere Wechselwirkungen in der aktiven Tasche koordiniert. Das an der Bindung des Diphosphatrestes beteiligte  $Mg^{2+}$ -Ion (Abb. 4; rot umrahmt) wird zusätzlich durch ein Wassermolekül koordiniert, das vermutlich während der Übertragung durch die Hydroxyl-Gruppe des konservierten Serins besetzt wird[36]. Die 3'-Phosphatgruppe (Abb. 4; orange umrahmt) bildet nur geringfügige Interaktionen mit dem Protein aus. Der  $\beta$ -Mercaptoethylamin-Rest mit seiner terminalen Thiolgruppe (Abb. 4; blau umrahmt) geht dagegen keine Interaktionen mit dem Protein ein. Daher sind auch chemische Modifikationen der zu übertragenden CoA-Derivate möglich[37]. In der letzten Gruppe sind die PPTasen vertreten, die in multifunktionale Enzyme am C-Terminus integriert sind und somit *in cis* agieren, wie bereits für *S. cerevisiae* für Typ II FAS gezeigt wurde[38].

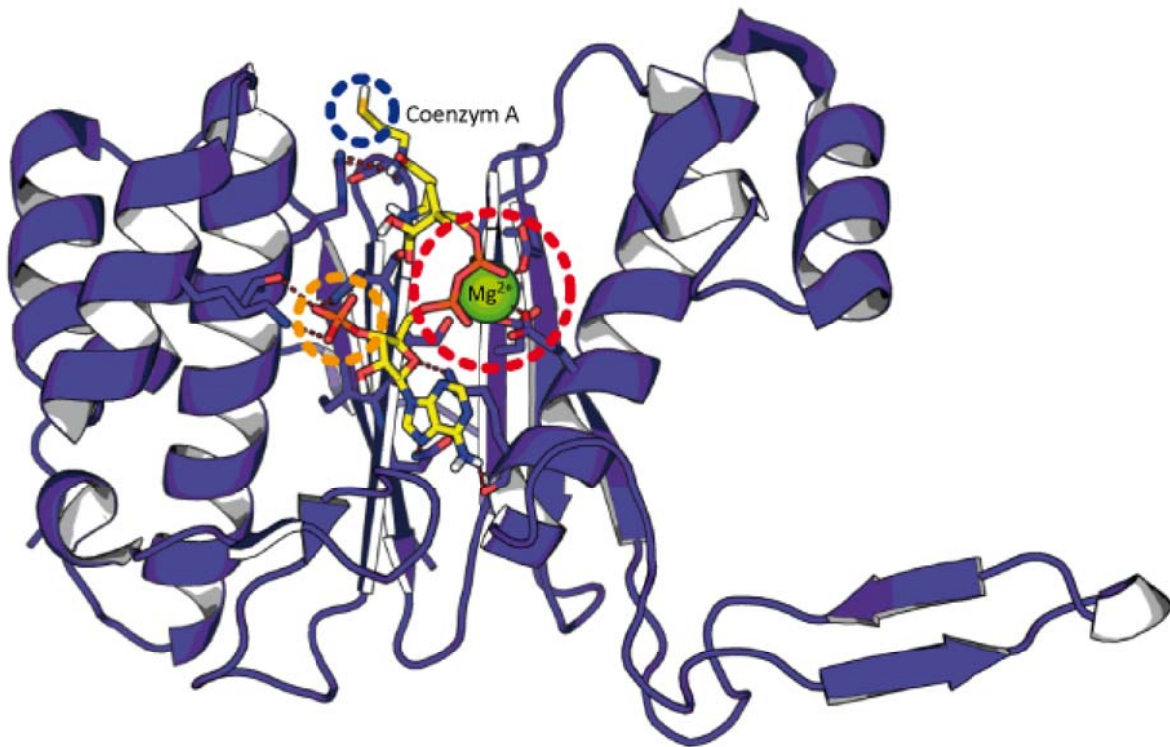


Abbildung 4: Phosphopantetheinyl-Transferase aus *B. subtilis* (Sfp; pdb: 1QR0).

## 1.2.1 Fettsäuresynthase (FAS)

Unverzichtbar für alle Organismen katalysieren Fettsäuresynthasen (FAS) die Biosynthese von Fettsäuren über einen iterativen Prozess, bei dem die Kohlenstoffkette der Fettsäure durch die wiederholte Kondensation mit dem ACP-gebundenem Malonat elongiert wird[39]. Die

resultierenden  $\beta$ -Keto-Thioester werden durch die Folge von Ketoreduktion, Dehydration und Enoylreduktion zu vollständig gesättigten Acyl-Thioestern umgewandelt (Abb. 5A). Jeder Organismus verfügt über ein Set an konservierten Enzymen, um die Katalyse der einzelnen Schritte der *de novo* Fettsäurebiosynthese zu übernehmen, obwohl das Zusammenspiel und die Architektur dieser Enzyme stark voneinander abweichen können. Bakterielle FAS (Typ II) [40-41] bestehen aus diskreten Proteinen, während bei Säugetieren und Pilzen (Typ I)[42] alle einzelnen enzymatischen Aktivitäten an einer Polypeptidkette stattfinden. Die Fettsäuresynthase (FAS) des Typ I wird in zwei weitere Subtypen eingeteilt, die Vertreter des Typ Ia finden sich in Pilzen sowie in wenigen Bakterien und bestehen entweder aus  $\alpha 6\beta 6$  (2.6 MDa) oder  $\alpha 6$  (1.3 MDa) Einheiten (fFAS; Abb. 5B) während die FAS in Säugetieren den Subtyp Ib vertreten und aus einem  $\alpha 2$ -Dimer (540 kDa) bestehen (mFAS; Abb. 5C)[43].

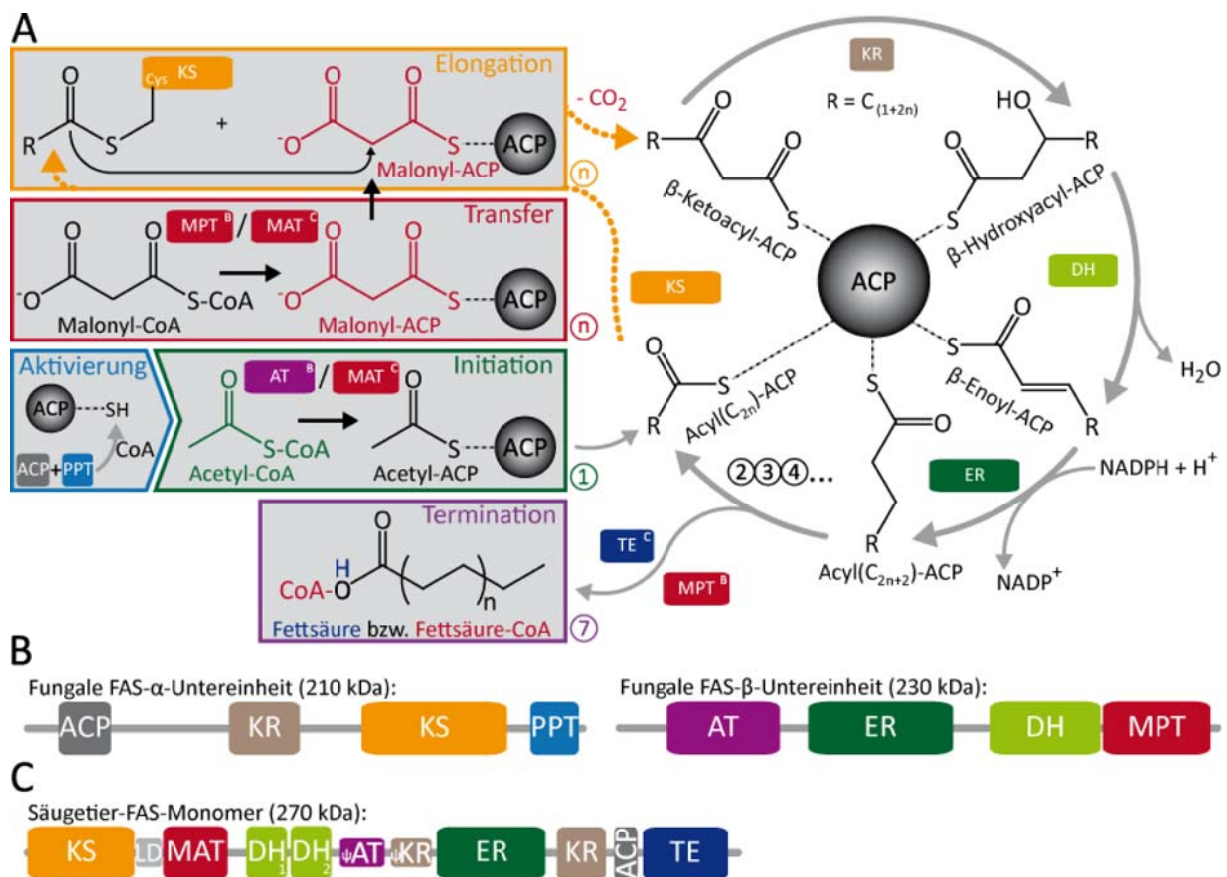


Abbildung 5: Domänenorganisation und Reaktionsschema der Fettsäurebiosynthese Typ I. (A) Reaktionsschema der Fettsäurebiosynthese (Lipogenese) in eukaryotischen Zellen. (B) Domänenorganisation in Pilzen (fungale FAS; fFAS). (C) Domänenorganisation in Säugetieren (mFAS). Abgebildete Domänen: Linker-Domäne (LD), Acyl-Carrier-Protein (ACP),  $\beta$ -Ketoreduktase (KR), Ketoacyl-Synthase (KS), Phosphopantethein-Transferase (PPT), Acetyl-Transferase (AT), Enoyl-Reduktase (ER), Dehydratase (DH), Malonyl-Palmitoyl-Transferase (MPT), Malonyl-Acetyl-Transferase (MAT) und Thioesterase (TE).

In allen Typen spielt das ACP eine zentrale Rolle und muss zunächst mittels eines posttranslationalen Phosphopantetheinyl-Transfers durch eine AcpS-ähnliche PPTase (Typ II) oder eine Sfp-ähnliche PPTase bei Säugetieren (Typ Ib) bzw. eine in die Polypeptidkette integrierte PPTase bei Pilzen (Typ Ia; Abb. 5B) in die aktive Holo-Form überführt werden[44]. In beiden Haupttypen verläuft die Initiation der Biosynthese über die Acylierung des Holo-ACP durch eine Acetyltransferase (AT) unter Verbrauch von Acetyl-CoA (Typ I) und durch die Acetyl-CoA-Carboxylase (ACC) mit Carbonat als Substrat (Typ II). Die Initiationsreaktion der Typ II FAS kann alternativ auch durch die Autoacylierung[45] sowie Automalonylierung[46] der ACP erfolgen. Dabei wird das Acetyl-CoA bzw. Malonyl-CoA von holo-ACP koordiniert, wodurch die saure Seitenkette eines Glutamatrestes von Helix III mit der Säuregruppe des jeweiligen Substrates interagiert. Danach findet ein nukleophiler Angriff der ACP-gebundenen Thiolgruppe des 4'-Phosphopantetheinyl-Armes (PPant) am Thioester-Kohlenstoff des jeweiligen Substrates statt. Auch andere Substrate wie Oxalsäure oder Thioester der  $\beta$ -Ketosäuren können in dem Autokatalytischen Prozess übertragen werden[45]. Durch die Malonyl/Palmitoyl-Transferase (MPT) bei Typ Ia bzw. Malonyl-Acetyl-Transferase (MAT) bei Typ Ib (FabD bei Typ II) findet die Übertragung des Malonyl-Restes auf das Holo-ACP statt. Durch die  $\beta$ -Ketoacylsynthase (KS) wird die Kondensationsreaktion mit dem Acetyl-CoA aus dem Initiationsprozess katalysiert und die Fettsäurebiosynthese (Liponeogenese) wird eingeleitet. In den drei folgenden Schritten wird zuerst in einer NADPH-abhängigen Reaktion das  $\beta$ -Ketosäure-Derivat, das erste Substrat in dem FAS-Zyklus, durch die Ketoacylreduktase (KR) reduziert. Im folgenden Schritt wird die so eingeführte Hydroxylgruppe durch eine (Typ Ia) oder zwei (Typ Ib und Typ II) Dehydratase-Domänen (DH) abgespalten und eine Trans-2-Enoyl-Gruppe wird gebildet. Im letzten Schritt des FAS-Zyklus wird die Doppelbindung des Enoyl-Derivates in einer NADPH- bzw. NADH-abhängigen Reaktion (Typ I bzw. Typ II), katalysiert durch die Enoylreduktase (ER), reduziert. Sobald die wachsende Fettsäurekette die entsprechende Länge erreicht hat, wird sie mittels der MPT-Domäne (Typ Ia) bzw. der Thioesterase (TE) bei FAS Typ Ib und II vom ACP abgespalten[47].

### *i.) Architektur eukaryontischer Fettsäuresynthasen in Säugetieren*

Auch wenn der gemeinsame Reaktionszyklus und die strukturelle Erhaltung der beteiligten katalytischen Domänen der eukaryotischen Fettsäuresynthasen (FAS Typ Ia und Ib) ver-

gleichbar sind, zeigen sich auf den verschiedenen Ebenen der Organisation der Fettsäuresynthasen in Säugetieren und Pilzen von der primären bis zu quartären Struktur signifikante Unterschiede. Bei der FAS in Säugetieren (Typ Ib) sind alle funktionalen Domänen in einer einzigen, ca. 2500 Aminosäuren langen Polypeptidkette (Abb. 5C) organisiert. Die katalytische Aktivität kann sich jedoch erst nach der Homo-Dimerisierung entfalten. Die aktive Form (540 kDa) bildet eine X-Form[48] bestehend aus dem, für die Kondensation zuständigen Bereich, der die MAT- und KS-Domänen beinhaltet, sowie dem, für die  $\beta$ -Kohlenstoff-Modifikationen zuständigen Bereich, der die DH-, KR- und ER-Domänen beinhaltet (Abb. 6A)[48].

Der für die Modifikationen zuständige Bereich enthält außerdem zwei weitere, katalytisch nicht aktive Domänen, die nach ihrer strukturellen Ähnlichkeit zu den aktiven KR- bzw. ME-Domänen als Pseudo-Ketoreduktase ( $\psi$ -KR) und Pseudo-Methyltransferase ( $\psi$ -ME) bezeichnet werden (Abb. 6B). Die zwei in der X-Form dimerisierten Polypeptidketten erzeugen zwei äquatorial angeordnete Reaktionskammern (*chamber*), wobei um jede Kammer alle für die Fettsäure-Kettenelongation benötigten katalytischen Zentren angeordnet sind. Die terminalen, durch eine flexible Prolin- und Alanin-reiche Linkerregion verbundenen Bereiche bestehend aus ACP- und TE-Domänen[49] können aufgrund der fehlenden Elektronendichte nicht visualisiert werden. Ihre Strukturen wurden als einzelne isolierte Polypeptidketten in hoher Auflösung gelöst[50-51]. Während die Faltung der KS-, KR- und der MAT-Domäne der ihrer bakteriellen Pendanten entsprechen und die DH-Domäne als Pseudo-Dimer vorliegt, vergleichbar mit dem bakteriellen DH-Homo-Dimer, nehmen die ER- und TE-Domäne eine für die Säugetier-FAS charakteristische Faltung an[49]. Während eines Zyklus interagiert die ACP-Domäne mit der MAT-Domäne und es erfolgt die initiale Beladung des PPant-Armes (Abb. 6C). Dafür neigt sich der untere Bereich um 25° in Richtung des Modifikationsbereiches und bewegt die MAT-Domäne in den ca. 75 Å langen Wirkungsradius der Holo-ACP-Domäne (*close chamber*). Der Acetyl-Rest wird zuerst auf die KS-Domäne der gleichen Polypeptidkette übertragen, so dass durch die MAT-Domäne der zweiten Polypeptidkette ein Malonyl-Rest an die ACP-Domäne gebunden werden kann. Das Substrat wird anschließend durch die acetylierte KS-Domäne elongiert.



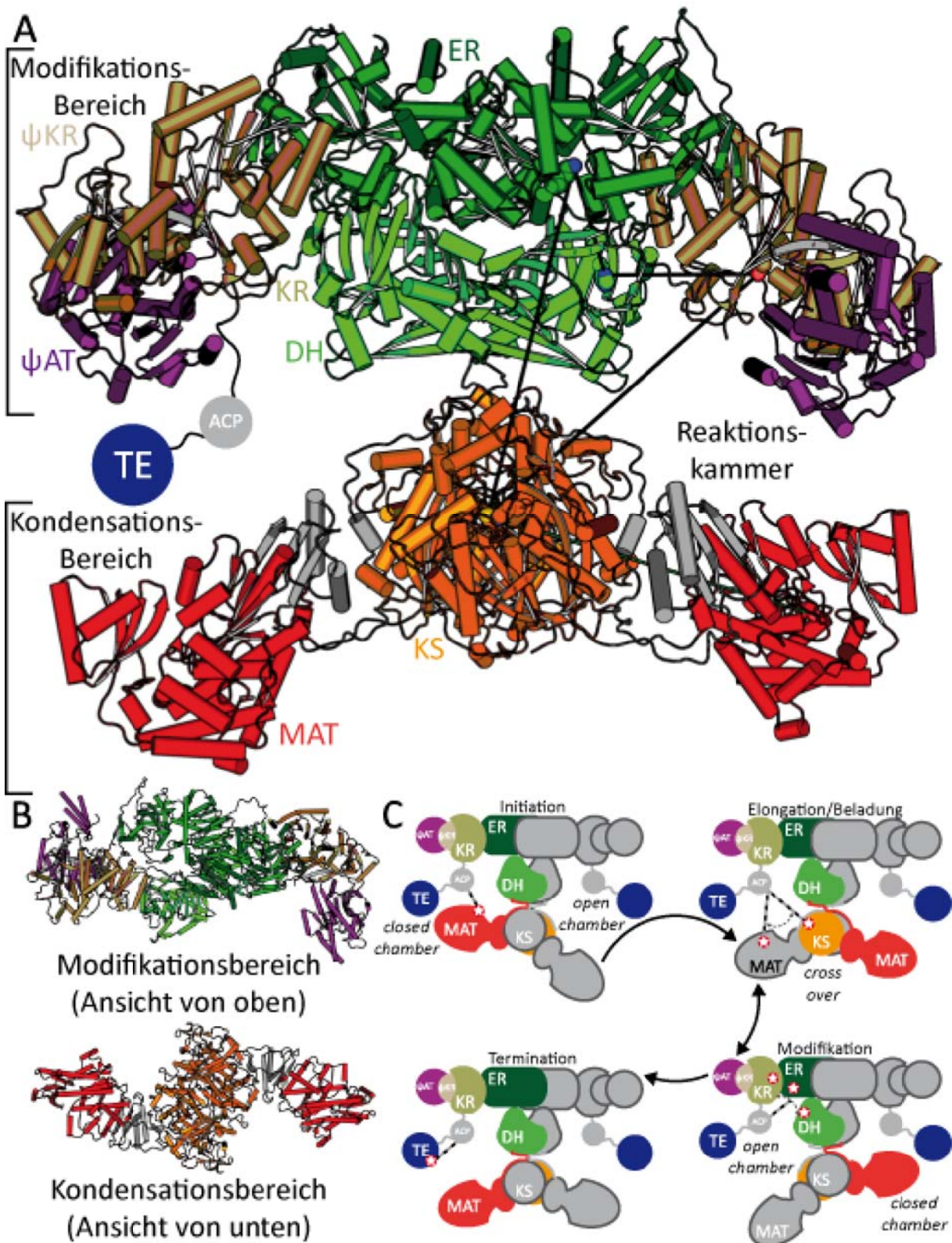


Abbildung 6: Strukturmodell und Schema der Fettsäuresynthese in Säugetieren. (A) Kristallstruktur der FAS Typ Ib (pdb: 2VZ9), der bewegliche terminale Bereich ist aufgrund fehlender Elektronendichte schematisiert. (B) Der Modifikations- und Kondensationsbereich. (C) Schematische Darstellung des FAS-Zyklus.

Durch die geschlossene Konstellation der einen Seite des Dimeres, sind die aktiven Zentren der DH- und ER-Domänen im oberen Modifikationsbereich der gegenüberliegenden Seite (*open chamber*) durch die ACP-Domäne besser zugänglich. Für die Kettenelongation schwenkt der sehr flexible untere Kondensationsbereich in die aktive Stellung und dreht sich dafür um 180° am Schwerpunkt des Dimeres (*cross over*)[48, 52]. Die Architektur der strukturell eng verwandten Polyketidsynthasen ermöglicht ein breites Spektrum an bioaktiven sekundären Metaboliten[53].

## ii.) Architektur eukaryontischer Fettsäuresynthasen aus Pilzen

Bei der Fettsäuresynthase aus Pilzen (Typ Ia), einem  $\alpha_6\beta_6$  Heterododekamer, sind im Gegensatz zu der FAS aus Säugtieren (Typ Ib) die katalytisch aktiven Domänen über die zwei Polypeptidketten geordnet (Abb. 5B-C). Die Typ Ia FAS nimmt eine tonnenförmige Struktur (*barrel shape*; Abb. 7A) ein, die aus einem durch sechs  $\alpha$ -Untereinheiten gebildeten zentralen Rad mit zwei (oberhalb und unterhalb) umgebenden Kuppeln aus jeweils drei  $\beta$ -Untereinheiten besteht (Abb. 7B-C). Die Kuppeln umschließen zwei separate Reaktionskammern, die jeweils drei Sets enzymatischer Domänen beinhalten, die für die Fettsäuresynthese durch die in jeder Kammer vorhandenen ACP-Domänen verantwortlich sind[54]. Die elektrostatisch negativ geladenen ACP-Domänen werden vor der Ausbildung des Heterododekamer durch die PPTase, in der Kristallstruktur lediglich durch den Linker vertreten (Abb. 7D), mit dem Ppant-Arm beladen[55] und interagieren in einer der sechs Kammern des Heterododekamer mit den sechs verschiedenen, positiv geladenen katalytischen Domänen[54] in einer Kreisbahn. Die ACP-Domäne ist beidseitig mit dem Gerüst der FAS verbunden. Der N-Terminus ist mit der Seitenwand der Reaktionskammer (peripherer Ankerpunkt) und der C-Terminus mit der Mitte des zentralen Rades (zentraler Ankerpunkt) verbunden (Abb. 7D). Der Abstand unter den einzelnen zentralen Ankerpunkten beträgt ca. 55 Å und zwischen den peripheren Ankerpunkten ca. 85 Å. Der N-terminale Linker der ACP-Domäne ist Prolin- und Alanin-reich und ermöglicht die uneingeschränkte Beweglichkeit der ACP-Domäne[54]. Nach einer initiativen Beladung der ACP durch die AT-Domäne ( $\alpha/\beta$ -Faltung) wird das für den nächsten Elongationsschritt benötigte Produkt in der KS-Domäne (Thiolase-Faltung) gebunden.



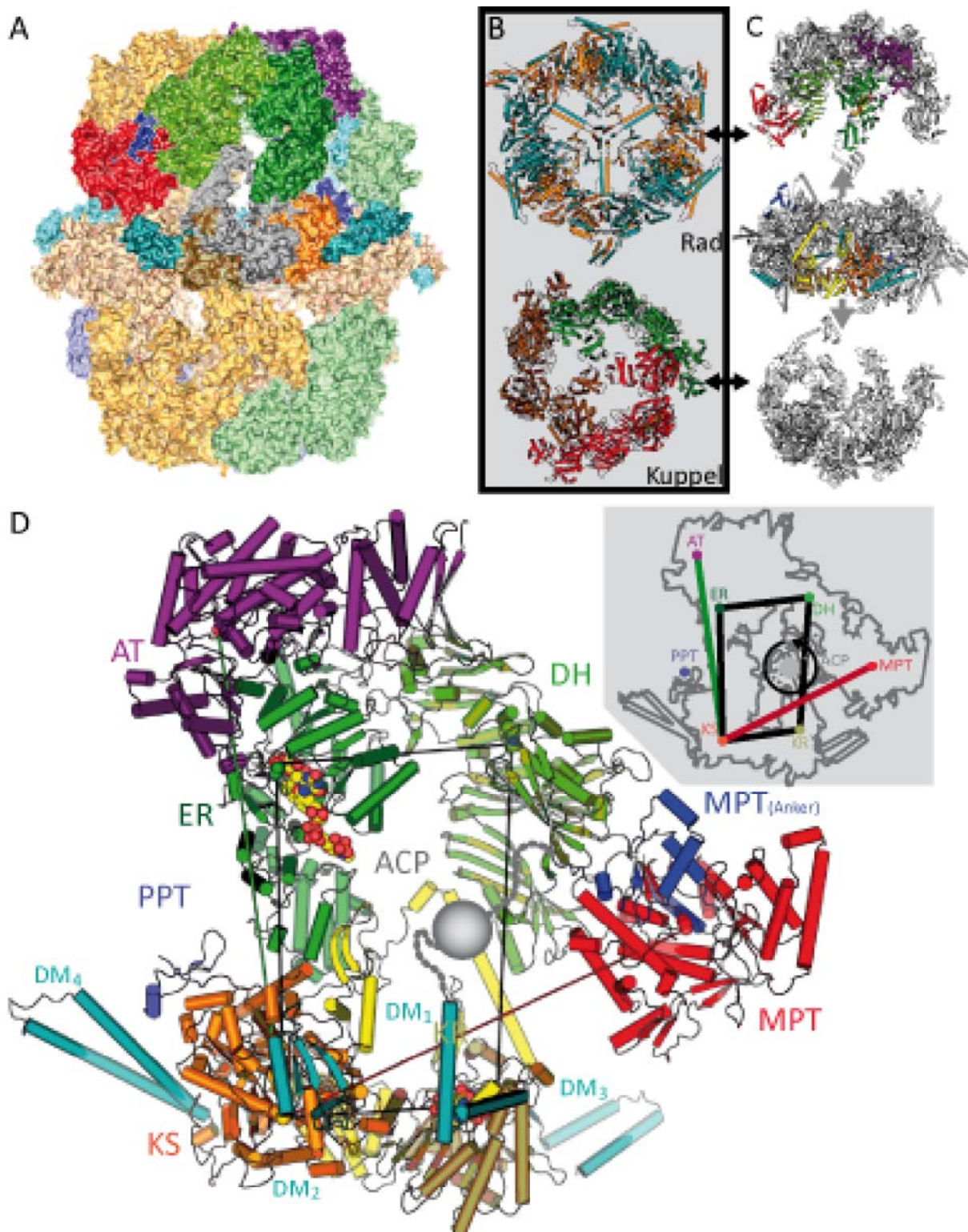


Abbildung 7: Strukturmodell des  $\alpha_6\beta_6$  Heterododecamer der Fettsäuresynthase aus Pilzen (A) Barrel-Struktur der FAS, einzelne Untereinheiten sind farblich markiert, der frontale Funktionsdimer ist nach Domänen farblich markiert. (B) Die Struktur des zentralen Rades (pdb: 2UVB) und der oberen und unteren Kuppel (pdb: 2UVC). (C) Die Struktur des zentralen Rades und der Kuppel von oben. (D) Das Zusammenspiel der  $\alpha$ - und  $\beta$ -Untereinheit und Ausbildung einer Kammer mit allen katalytischen Domänen sowie der Reaktionsabfolge.



Anschließend erfolgt eine Beladung durch die MPT-Domäne ( $\alpha/\beta$ -Faltung), bevor die Modifikationen durch die NADPH-abhängige KR- (Rossmann-Faltung), DH- (dreifache Hot-Dog-Faltung) und die NADPH- und FMNH<sub>2</sub>-abhängige ER-Domäne (TIM-Barrel-Faltung) erfolgen. Die Elongation erfolgt durch die Kondensation mit dem an der KS-Domäne koordinierten Zwischenprodukt (Abb 7D). Die für die Fettsäuresynthese benötigten Substrate können durch die seitlichen Öffnungen in die einzelnen Kammern diffundieren[54]. Während bei der FAS aus Säugetieren nur 9% der Aminosäuren (die nicht aktiven Domänen nicht einbezogen) die Linkerregionen ausbilden[43], definieren bei den Pilzen ca. 50% der Aminosäuren das starre, aus den Dimerisierungsdomänen (DM), Insertionen bestehende (Abb. 7D; cyan und gelb) Gerüst der FAS[54].

## 1.2.2 Polyketidsynthase (PKS)

Polyketide stellen eine sehr heterogene Gruppe von Naturprodukten mit strukturell anspruchsvollen Eigenschaften dar, die aus einem Kohlenstoffgerüst aufgebaut sind[56]. Viele Naturstoffgruppen wie Polyphenole, Makrolide, Polyene, Endiine oder Polyether werden durch Polyketidsynthasen (PKS) synthetisiert[57]. Die Naturstoffe erfüllen dabei verschiedene Funktionen als Virulenzfaktoren, Pigmente oder Verteidigungs- und Kommunikationsstoffe. Trotz der hohen strukturellen und funktionellen Vielfalt werden die Polyketide nur aus zwei Bausteinen, dem Acetat und Propionat zusammengesetzt. Polyketidsynthasen zeigen viele Homologien, sowohl bei dem Substratpool (Acetyl-CoA und Malonyl-CoA), als auch im funktionellen und strukturellen Aufbau. Sowohl PKS als auch mFAS verwenden die decarboxylative Claisen-Thioester-Kondensation durch die  $\beta$ -Ketosynthase (KS). Trotz der vielen Homologien sind Modifikationen an der wachsenden Produktkette bei PKS-Systemen nur optional, dagegen sind sie in FAS-Systemen an jedem Zyklus beteiligt. Während der Polyketid-Synthese greift die Seitenkette des konservierten Serins der Acyltransferase (AT) nukleophil den Malonyl-CoA-Thioester an und bildet ein Acyl-O-Enzym-Intermediat. Anschließend wird der Malonylrest auf die Thiolgruppe des ACP-gebundenen PPant-Armes übertragen. Die folgende Kondensationsreaktion, die nach der Translokation des Zwischenproduktes und der Decarboxylierung des abwärtsliegenden (*downstream*) Malonyl-S-ACP stattfindet, wird durch die Ketosynthase (KS) katalysiert[58]. Die resultierende

Keto-Kette kann dann durch die Enzyme Ketoreduktase (KR)[59], Dehydratase (DH)[60] und Enoylreduktase (ER)[61] modifiziert werden. Durch ein FAS-System wird eine bestimmte Fettsäure produziert, während bei den PKS-Systemen alternative Produkte wie Olefine, Alkohole oder Ketone gebildet werden können[62]. Die Abspaltung des Produktes vom ACP-gebundenem P<sub>ant</sub>-Arm wird durch die Thioesterase (TE) katalysiert. Die strukturelle Organisation der Polyketidsynthasen kann ähnlich der Klassifizierung bei Fettsäuresynthasen in diverse Klassen unterteilt werden. Die PKS-Systeme des Typs I sind sehr große lineare (6-Deoxyerythronolid B[63]) oder iterative (Lovestatin[64]) multifunktionale Enzyme, wohingegen PKS-Systeme des Typs II konkrete, einzelne, monofunktionale Enzyme[65] enthalten. Die PKS-Systeme des Typs III werden zu den Chalkon-Synthasen gezählt[66].

### *i.) Architektur der Polyketidsynthasen*

Zu den bestuntersuchten PKS-Systemen gehören die drei 6-Deoxyerythronolid-B-Synthasen (DEBS)[67]. Der 194 kDa schwere Homodimer (KS-AT) des DEBS-Modul 5 ist die erste strukturell aufgeklärte Bidomäne der PKS-Systeme (pdb: 2HG4). Das kristallisierte Fragment beinhaltet die vollständig ausgebildete Struktur der Ketosynthase (KS) und Acetyltransferase (AT) sowie drei Linkerbereiche (Abb. 8A): die N-terminale Dimerisierungsregion (gelb), den interdomenalen KS-AT-Linker (grau) und den C-terminalen Peptidlinker (blau). Auch wenn die Ketoreduktase (KR) sowie die ACP-Domäne des Modul 5 (Abb. 8B) in der Kristallstruktur nicht vertreten sind, stellt mit der Domänenkonstellation KS-AT diese Struktur 95% eines Minimalmoduls dar. Das Minimalmodul besteht ausschließlich aus den für die Produktelongation essentiellen KS-, AT- und ACP-Domänen, wie bereits unter anderem für das Modul 3 und 6 der Narboldid Synthase[68] oder Modul 2 der Rifamycin Synthase[69] festgestellt. Zusätzlich wurden die Kenntnisse über die Architektur der PKS-Systeme durch die strukturellen Untersuchungen der KS-AT-Bidomäne aus Modul 3 der 6-Deoxyerythronolid-B-Synthase (pdb: 2QO3)[70] erweitert. Im DEBS Modul 5 werden Dimergrenzfläche (*interface*) durch die Protein-Protein-Interaktionen sowohl durch das KS-Domänen-Dimer als auch durch die *Coiled-Coil*-Struktur des N-terminalen Linkers ausgebildet (Abb. 8A). Die KS-Domänen weisen eine  $\alpha\beta\alpha\beta\alpha\beta$ -Faltung auf, die mit den Kondensations-Enzymen der Typ II FAS-[71] und PKS-Systeme vergleichbar ist.

Die Dimerisierung der KS-Domänen erfolgt sowohl über die Rückgrat-Interaktion zweier  $\beta$ -Faltblätter (AA 192-196) als auch über den *Coiled-Coil*-Linker am N-Terminus, dessen  $\alpha$ -Helices über eine hydrophobe Grenzfläche interagieren. Die Acetyltransferase zeigt eine  $\alpha$ , $\beta$ -Hydrolase-ähnliche Faltung im Kern sowie eine Ferredoxin-Faltung der Unterdomäne. Eine neuartige Faltung weist die Linkerregion zwischen der KS- und AT-Domäne auf, die aus drei  $\beta$ -Faltblättern und zwei  $\alpha$ -Helices besteht. Die beiden  $\alpha$ -Helices interagieren mit der  $\alpha$ -Helix des C-terminalen Peptidlinkers und formen eine  $\alpha\beta\alpha\beta$ -Faltung. Der restliche Teil des Peptidlinkers (30 AA) interagiert mit der KS-Domäne. Die im aktiven Zentrum konservierten Aminosäuren des KS-AT-Dimers, Cystein<sup>199</sup> in der KS<sup>1</sup>- und Serin<sup>642</sup> in der AT<sup>2</sup>-Domäne, liegen ca. 80 Å voneinander entfernt. Die in der Elektronendichte nicht aufgelöste ACP-Domäne muss in einem Minimalsystem dynamisch mit beiden Domänen abwechselnd interagieren, da der Bewegungsradius der prosthetischen Gruppe alleine nicht ausreichend ist, um das

gebundene Substrat in die Nähe der beiden aktiven Zentren zu bringen. In weiteren Untersuchungen wurden die Interaktionen der ACP- und der KS-Domäne durch eine Modellierung über die *Rigid-Body*-Methode simuliert. Gleichzeitig wurden verschiedene Chimere der ACP-Domäne konstruiert und auf ihre Fähigkeit mit dem bidomänen KS-AT-System zu interagieren hin getestet. Mit der so ermittelten ACP-Interaktionsfläche wurde die Kontaktregion an dem KS-AT-Linkerbereich (Abb. 8B, grau) identifiziert[72]. Die aus der Interaktionsfurche heraus mit den aktiven Zentren der KS<sup>2</sup>- und AT<sup>1</sup>-Domäne agierende ACP-Domäne benötigt die Dimer-Anordnung für die entsprechende katalytische Aktivität. Somit wurde der Grundstein für das Verständnis der katalytischen Aktivität von Polyketidsynthasen als Dimer gelegt, sowie ein genereller Ansatz für den Unterschied zwischen der Architektur der als Dimer agierenden PKS- und der als Monomer agierenden nichtribosomalen Peptidsynthetasen.

### 1.2.3 Nichtribosomale Peptidsynthetase (NRPS)

Während bei den Fettsäuresynthasen die Struktur und Länge der elongierten, ACP-gebundenen Kette durch die enzymatische Katalyse der Elongations- und Modifikationsenzyme bestimmt wird, folgt die Polyketidsynthase einer modularen Syntheselogik. Durch diese ist es möglich eine wachsende Kette bestehend aus verschiedenen Bausteinen, die durch die korrespondierenden Module aktiviert, kondensiert und modifiziert werden, darzustellen. Der gleichen Syntheselogik folgt auch die nichtribosomale Peptidsynthetase (NRPS). Diese modular aufgebauten Multienzymkomplexe bestehen aus einzelnen, strukturell definierten Untereinheiten, den Domänen, die für die Aktivierung, den Transport, die Elongation und die Modifikation der korrespondierenden Aminosäuren und Zwischenprodukte sowie für die Termination der Peptidsynthese verantwortlich sind[73]. Dabei entspricht die Reihenfolge der Peptidbausteine im Produkt der Anordnung der entsprechenden Module, die aus Wiederholungen der Kondensations- (C), Adenylierungs- (A) und Peptidyl-Carrier-Protein-Domäne (PCP; T) sowie eventuellen Modifikationsdomänen bestehen (NRPS Typ A; Abb. 9). Aufgrund der räumlichen Verteilung der aktiven Zentren einzelner Domänen werden die kovalent gebundenen Substrate, vergleichbar mit einem Fließband, durch die dynamische PCP-Domäne von einem aktiven Zentrum zum nächsten

transportiert[74]. Analog zu der Rolle der Holo-ACP-Domäne in FAS- und PKS-Systemen, ist für den Transport der wachsenden Peptidkette bei NRPS die Holo-PCP-Domäne (*Peptidyl Carrier Protein*) verantwortlich[75]. Durch den Bewegungsradius der dynamischen PCP-Domäne können dabei Distanzen überwunden werden, die über die maximale Reichweite von 20 Å des kovalent gebundenen prosthetischen Armes hinausgehen. Innerhalb eines Modules wird die korrespondierende Aminosäure durch die Adenylierungsdomäne (A-Domäne) unter Verbrauch von ATP aktiviert (Adenylierung) und anschließend als Thioester auf den PCP-gebundenen PPant-Arm übertragen (Thiolierung). Ein Minimalmodul wird durch die PCP- und A-Domäne definiert, für Verknüpfungen weiterer Aminosäuren wird die Kondensationsdomäne (C-Domäne) benötigt. Diese katalysiert die Kondensationsreaktion zwischen der an der Donor-PCP (*Upstream-PCP*) wachsenden Peptidkette und der an der Akzeptor-PCP (*Downstream-PCP*) gebundenen Aminosäure. Die Synthese eines nicht-ribosomalen Peptides beginnt am ersten Modul eines NRPS-Systems (Initialmodul). Die Initiation erfolgt dabei entweder durch die Aktivierung der ersten Aminosäure (Initiationsmodul: A-PCP) wie bei der Tyrocidin A-Synthetase[76], oder durch Kondensation mit einem carboxylierten Rest (Initiationsmodul: C-A-PCP), wie bei der Kondensation einer  $\beta$ -Hydroxy-Fettsäure bei der Surfactin Synthetase SrfAA[77] oder der Kupplung von PCP-gebundenem 2,3-Dihydroxybenzoat (DHB) mit dem modulgebundenem Serin bei der Enterobactin Synthetase EntF (EntF-Modul: C-A-PCP-TE)[78]. Desweiteren kann die Initiation durch Kupplung einer am N-Terminus formylierten Aminosäure wie bei der LgrA-Synthetase[79] erfolgen, durch folgende Elongationsmodule verlängert und am Terminationsmodul auf die Thioesterase (TE-Domäne) übertragen werden. Dort wird sie hydrolytisch oder zyklisch abgespalten. Die Synthese der nichtribosomalen Peptide ist nur mit Hilfe dieser Domänen möglich, sie sind somit essentiell (Abb. 9-1). Zusätzlich wird die Zahl der möglichen Bausteine durch viele Modifikationsreaktionen auf bislang ca. 250 erweitert[80]. Bezieht man auch Hybridsysteme mit ein (PKS-NRPS, glycosylierte Bausteine, Fettsäuren) konnten bislang über 500 verschiedene Bausteine für die nichtribosomalen Peptiden identifiziert werden[81]. Die Modifikationen können durch die Substitution essentieller Domänen mit einer alternativen Domäne erreicht werden.

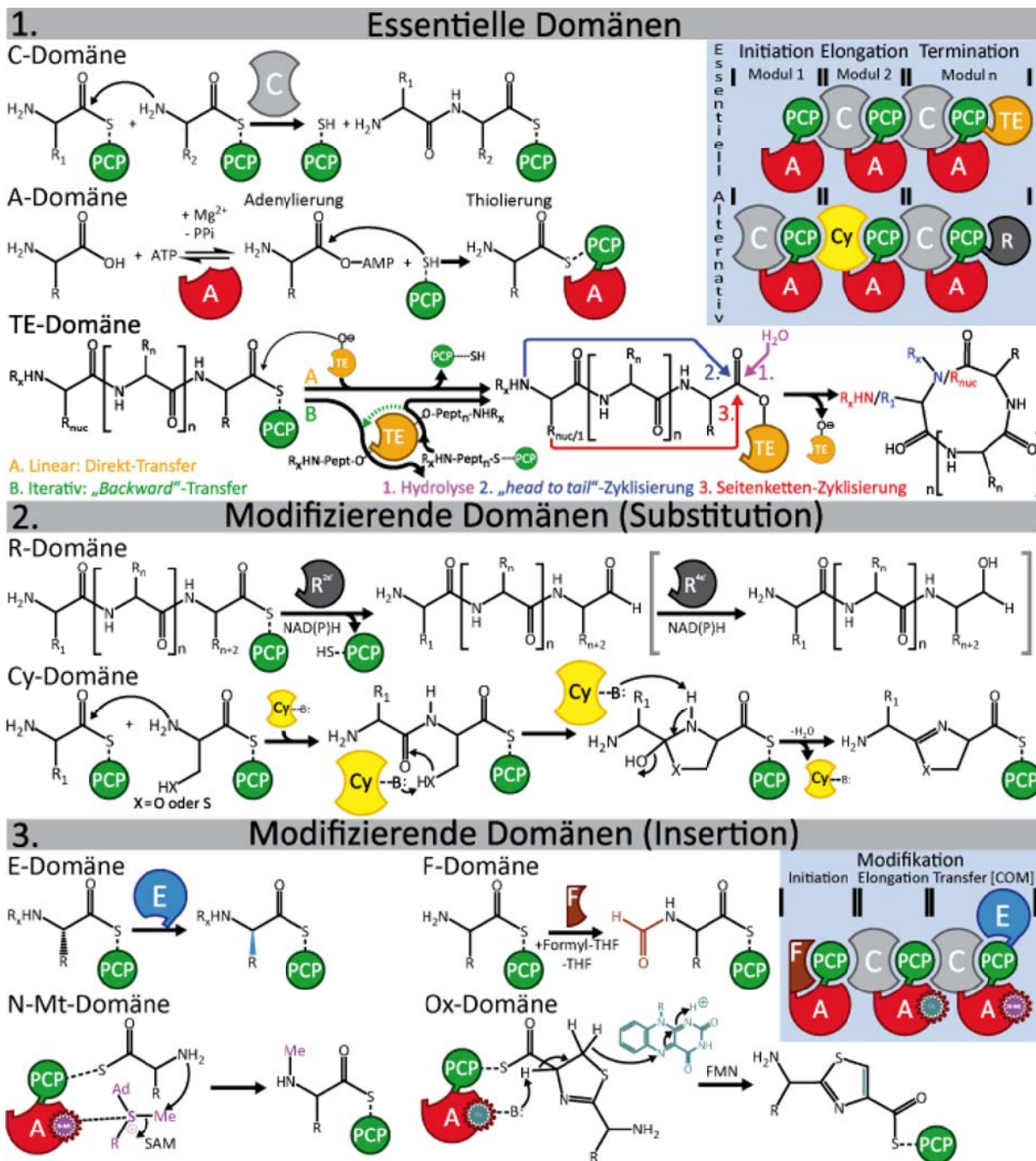


Abbildung 9: Katalysemechanismen und Aufbau der nichtribosomalen Peptidsynthetasen. Schematische Darstellung der von essentiellen (1.) sowie modifizierenden (2. und 3.) Domänen katalysierten Reaktionen sowie der repräsentativen, modular aufgebauten, (NRPS). Abgebildete Domänen: Kondensationsdomäne (C), Adenylierungsdomäne (A), Thioesterase (TE), Reduktionsdomäne (R), Cyclisierungsdomäne (Cy), Epimerisierungsdomäne (E), Formylierungsdomäne (F), N-Methylierungsdomäne (N-Mt), Oxidationsdomäne (Ox).

Dabei erfolgt durch die Zyklisierungsdomäne[82] (Bacitracin Synthetase A[83]) sowohl die Heterozyklisierungs- als auch die Kondensationsreaktion, während die Reduktase (R-Domäne) des terminalen Moduls LgrD der linearen Gramicidin Synthetase[84] die



alternative reduktive Abspaltung des elongierten Peptides katalysiert (Abb. 9-2). Zusätzlich können durch die Insertion einzelner Domänen in die NRP-Synthetasestruktur, wie z.B. die N-terminale F-Domäne[79] oder die Epimerisierungsdomäne (E-Domäne)[85] Modifikation an den einzelnen Bausteinen durchgeführt werden. Desweiteren können Bausteine durch die in die A-Domäne integrierte Domänen wie die N-Methyltransferase (N-Mt-Domäne; Cyclosporin-Biosynthese[86]), sowie die Monooxidase (Ox-Domäne; Myxothiazol-Biosynthese[87]), modifiziert werden (Abb. 9-3). Außerdem können Modifikationsreaktionen sowohl vor der Synthese an den einzelnen Bausteinen, während der Synthese durch unabhängige Enzyme, oder auch nach der Synthese am fertigem Produkt erfolgen[88-89]. Durch die Sequenzanalyse einzelner Domänen konnten spezielle Bereiche jeder Domäne identifiziert werden, die konservierte Aminosäuresequenzen aufweisen. Diese sind für die enzymatische Aktivität, strukturelle Integrität sowie inter- und intradomänale Interaktionen zuständig. Diese Bereiche werden als Core-Motive (*core motif*) bezeichnet (Abb. 10)[90-92], die für die Faltung und Funktionalität der entsprechenden Domänen, trotz teilweise hoher Sequenz-Unterschiede, verantwortlich sind.

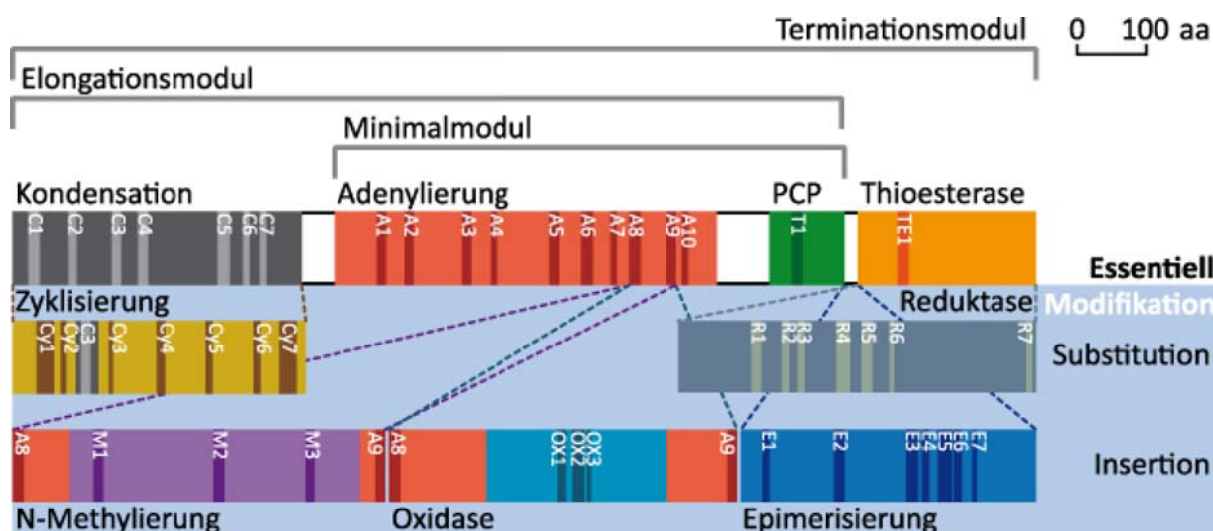


Abbildung 10: Verteilung und Position der NRPS-Domänen und deren Core-Motive.

Analog zu der Klassifizierung bei FAS- und PKS-Systemen, werden die verschiedenen NRPS-Systeme aufgrund der Reaktionsabfolge in drei unterschiedliche Typen eingeteilt[93]. Der Typ A beschreibt NRPS-Systeme, in denen die Anzahl und Reihenfolge der eingebauten Bausteine der Anzahl und Reihenfolge der A-Domänen entspricht. Ein linearer Modulaufbau und eine iterative Peptidelongation, bei der an der Thioesterase des terminalen Moduls

Peptidketten, die aus mehreren kürzeren Peptidketten bestehen, verknüpft werden beschreibt die Typ B NRPS Systeme (Thiocoralin-Biosynthese[94]). Nicht lineare NRPS-Systeme (Coelichelin-[95] oder Kutznerid-Biosynthese[96]) werden dem Typ C zugeordnet.

### *i.) Die Adenylierungsdomäne (A-Domäne)*

Die ca. 550 Aminosäuren lange Adenylierungsdomäne gehört neben den Acyl- und Aryl-CoA Synthetasen und der Luciferase zu der ANL-Superfamilie[97] (abgeleitet von Acyl-CoA-Synthetase (Acs), NRPS und Luciferase) der adenylierenden Enzyme. Alle Vertreter dieser Familie katalysieren zwei Teilreaktionen: die initiale Adenylierung des Substrat-Carboxylatrestes mit der Bildung eines Acyl-AMP-Zwischenproduktes und die anschließende Thiolierung durch die Katalyse zwischen der PPant-Thiolgruppe der PCP-Domäne und des Acyl-AMPs (Abb. 9-1). Obwohl viele Enzyme, wie z.B. die Aminoacyl-tRNA-Synthetase[98], ähnliche Katalysemechanismen aufweisen, zeigen sie keine Sequenz-Homologien zu den Mitgliedern der ANL-Superfamilie. Die ANL-Enzyme können zwar verschiedene Bausteine erkennen und umsetzen, jedoch weisen alle bei einer Sequenzidentität von ca. 20%, hohe strukturelle Homologien und den gleichen Katalysemechanismus auf.

Die Faltung der ANL-Enzyme zeigt zwei Untereinheiten (Abb. 11): eine größere N-terminale Core-Domäne (Aminosäurenreste 1-395) sowie eine kleinere C-terminale Sub-Domäne (Aminosäurenreste 400-550). Die Untereinheiten werden durch vier Aminosäuren (Scharnier) voneinander getrennt (Tab. 1, Core-Motiv A8). Die Core-Domäne besteht aus zwei Einheiten ähnlicher Topologie ( $\beta$ -Faltblätter flankiert durch  $\alpha$ -Helices) und einer dritten Einheit aus einer ungeordneten  $\beta$ -Barrel-Struktur. Die Sub-Domäne nimmt eine gemischte ( $\alpha$ + $\beta$ )-Struktur ein. Neben dem für die Funktion essentiellen flexiblen Scharnier ist der P-Loop (Tab. 1, Core-Motiv A3) das am höchsten konservierte Motiv der ANL-Enzyme[99]. Der Serin-, Threonin- und Glycin-reiche Bereich komplexiert die  $\beta,\gamma$ -Position des ATP-Phosphatrestes. Neben dem P-Loop sind zwei weitere Hauptmotive der ANL-Enzyme für die Koordination des ATP-Purin-Ringes sowie des  $Mg^{2+}$ -Ions (Kernmotiv A5 – Hauptmotiv II) und die Koordination der ATP-Ribose-Hydroxylgruppen (Kernmotiv A7 – Hauptmotiv III) verantwortlich[100].

Zwischen den Kernmotiven A4 und A5 liegt der variable Bereich, der für die richtige Erkennung der entsprechenden Aminosäure verantwortlich ist. Anhand dieses Bereiches ist es auf der Sequenzebene möglich, Voraussagen über die Spezifität der einzelnen A-Domänen



zu treffen[101]. Durch das erste Strukturmodell einer Luciferase (pdb: 1BA3)[102] konnten die durch Sequenzvergleiche und rationale Mutagenese bereits identifizierten Core-Motive (Tab. 1) ihrer strukturellen oder teilweise auch katalytischen Bedeutung zugeordnet werden. Die räumliche Anordnung des N- und C-terminalen Bereiches in der die Sub- von der Core-Domäne entfernt ist und keine Interaktionen stattfinden, entspricht der offenen Konformation (Abb. 11A).

Core-Motiv	Pos.	Sequenz	Funktion	
A-Dom.	A1	<u>Leu</u> 26	<u>L</u> (TS)YxEL	Strukturelement
	A2	<u>Arg</u> 74	L( <u>RKF</u> )AGxAYL(VL)P(LI)D	Strukturelement
	A3 (I)	<u>Tyr</u> 151	LAYxxY <u>T</u> SG(ST)TGxPKG	Katalyseelement ( <b>B</b> :P-Loop)
	A4	<u>Asp</u> 196	F <u>D</u> xS	Katalyseelement (AA-NH)
	A5 (II)	<u>Tyr</u> 296	NxYGP <u>T</u> E	Katalyseelement ( <b>B</b> :Purin, Mg <sup>2+</sup> )
	A6	<u>Gly</u> 351	GELxIx <u>G</u> x(VL)ARGYL	Strukturelement
	A7 (III)	<u>Asp</u> 382	Y(RK)TG <u>D</u> L	Katalyseelement ( <b>B</b> :Rib-OH)
	A8	<u>Asp</u> 389	GRx <u>D</u> xQVKIR <u>G</u> xRIELGEIE	<i>Scharnier</i> ( <u>D</u> ); Pantotheintunnel ( <b>C</b> : <u>G</u> )
	A9	<u>Pro</u> 470	LPxYM(IV)P	Strukturelement (NRPS)
	A10	<u>Lys</u> 491	NG <u>K</u> (VL)DR	Katalyseelement ( <b>B</b> )

Tabelle 1: Die Core-Motive der Adenylierungsdomänen sowie ihre Verteilung und Funktion

Alle bisher in der offenen Konformation kristallisierten ANL-Enzyme zeigen keine Inkorporation von ATP, Aminosäure oder CoA[103]. Das Strukturmodell von PheA (pdb: 1AMU), der Phenylalanin aktivierenden A-Domänen des Initiationsmodules aus dem Gramicidin S NRPS-System[104], weist die biochemisch bereits gezeigte Konformationsänderung der A-Domäne[98] auf und enthält das Substrat Phenylalanin sowie AMP (Abb. 11B). Bei PheA ist die C-terminale Sub-Domäne im Vergleich zu der Intermediat-Konformation geneigt ( $\Delta\phi = 100^\circ$ ). Erst in dieser Adenylierungs-Konformation sind Interaktionen zwischen den konservierten Core-Motiven und den Substraten auf der strukturellen Ebene sichtbar. Dabei wurde unter anderem die Interaktion des Lysinrestes aus dem A10-Motiv (Tab. 1) mit dem Ringsauerstoff der Ribose festgestellt. In weiteren Studien wurde die essentielle Bedeutung dieses Motives während der Adenylierung beobachtet[105]. Keine der Strukturen der ANL-Enzyme in der Adenylierungs-Konformation zeigte eine Interaktion mit

CoA oder einem vergleichbaren Thiolakzeptor.

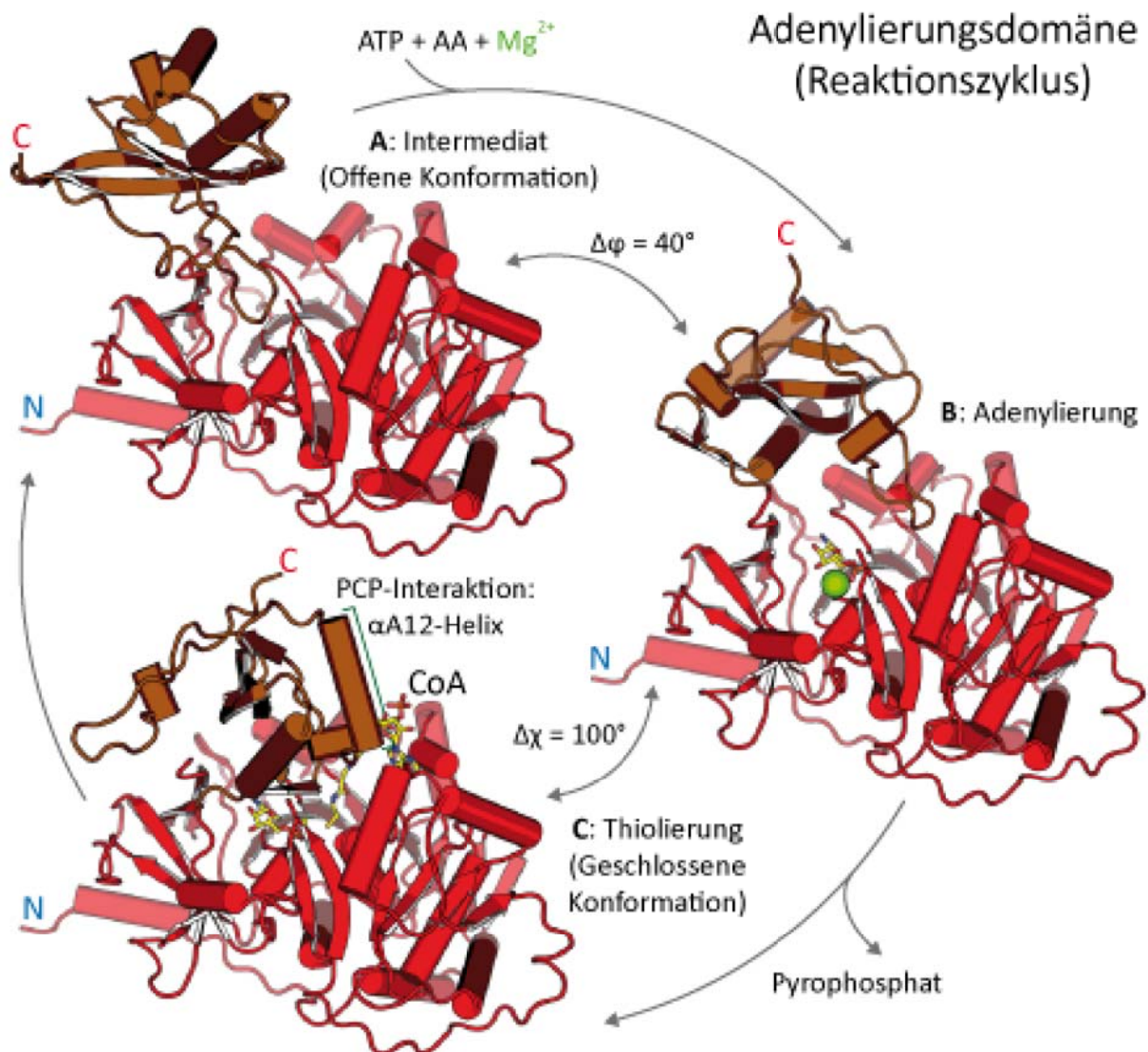


Abbildung 11: Reaktionszyklus der PheA A-Domäne und die Superposition der Sub-Domäne. (A) Intermediat: offene Konformation, Superposition der PheA Sub-Domäne gegen die Luciferase aus *P. pyralis* (pdb: 1BA3); (B) Adenylierung: A-Domäne des ersten GrsA Modules (PheA) aus *B. brevis* mit Phenylalanin, AMP und  $Mg^{2+}$  (pdb: 1AMU); (C) Thiolierung: geschlossene Konformation, Superposition der PheA Sub-Domäne mit der Acetyl-CoA-Synthetase (Acs) aus *S. enterica* mit CoA und AMP (pdb: 2P2F).

Für diese Interaktion ist eine Konformationsänderung notwendig[106]. Erst das Strukturmodell der Acs (pdb: 2P2F)[107] mit AMP und CoA zeigt die Interaktion des ANL-Enzymes mit dem Phosphopantethein-Rest des CoA. Die C-terminale Sub-Domäne weist dabei eine deutliche Neigung ( $\Delta\chi = 100^\circ$ ) im Vergleich zu der Adenylierungs-Konformation auf (Abb. 11C). Die  $\epsilon$ -Aminogruppe des Lysinrestes aus dem Core-Motiv A10 (Tab. 1) befindet sich dabei 25 Å von dem aktiven Zentrum entfernt. Außerdem dreht sich das durch die

Aminosäuren des Core-Motives A8 geformte  $\beta$ -Faltblatt zum aktiven Zentrum hin und bildet so den Pantetheintunnel aus. Basierend auf allen Daten wurde der Reaktionszyklus der ANL-Enzyme als alternierende Konformationsänderung beschrieben[107], während dem das Enzym die Adenylierungs-Konformation durch die Interaktion mit ATP,  $Mg^{2+}$ -Ion und dem korrespondierendem Substrat einnimmt (Abb. 11B). Dann findet die Adenylierungsreaktion sowie die damit verbundene Abspaltung des Pyrophosphats (PPi) statt. Durch eine anschließende Konformationsänderung bildet sich durch eine Neigung der Sub-Domäne und dem im Scharnier befindlichen  $\beta$ -Faltblatt (Tab. 1, A8) der Pantetheintunnel aus. Durch die nachfolgende Interaktion mit CoA oder Holo-PCP[108] wird das Substrat auf die entsprechende Thiolgruppe übertragen. Das Enzym zeigt dabei die geschlossene Thiolierungs-Konformation (Abb. 11C). Durch die erneute Neigung der Sub-Domäne von der Core-Domäne weg wird AMP freigesetzt und die A-Domäne nimmt so die Intermediat-Konformation (offen) ein (Abb. 11A). Anschließend kann der Reaktionszyklus erneut beginnen.

## ii.) Peptidyl Carrier Protein (PCP)

Die ca. 80 Aminosäuren langen Carrier-Proteine aus FAS-, PKS- (ACP) und NRPS-Systemen (PCP) und deren Hybridsystemen teilen eine gemeinsame Faltung bestehend aus drei bzw. vier  $\alpha$ -Helices, wie bereits durch die TycC<sub>3</sub>-PCP (pdb: 1DNY)[109] gezeigt wurde, sowie die gleiche Funktionalität mit der prosthetischen Gruppe (PPant-Arm). Der konservierte Serinrest (Tab. 2), an dem die prosthetische Gruppe kovalent gebunden ist, befindet sich am N-terminalen Ende der zweiten  $\alpha$ -Helix (Helix II).

Core-Motiv	Pos.	Sequenz	Funktion
<b>PCP</b>	T1	<u>Ser</u> 45 GG(H/D) <b>S</b> (L/I)	Katalyseelement: Pantetheinylierung

Tabelle 2: Der konservierte Serinrest der PCP-Domäne, Ankerpunkt des PPant-Armes

Alle FAS-, PKS- und NRPS-Systeme haben wie voneinander entfernte aktive Zentren, so dass der 20 Å Bewegungsradius der prosthetischen Gruppe nicht das erforderliche Spektrum an Beweglichkeit bietet um alle diese Zentren erreichen zu können. In NRPS-Systemen besteht die Funktion des Carrier-Proteins (PCP) darin gebundene Substrate in den Wirkungsbereich

weiterer aktiver Zentren zu transportieren. Daher ist die PCP-Domäne in den Adenylierungszyklus während der Substrataktivierung, den Kondensationszyklus während der Substratelongation sowie in die terminale Abspaltung des Produktes oder auch in eventuellen Modifikationsreaktionen involviert. Um ihre Aufgabe erfüllen zu können ist die dynamische PCP-Domäne in der Lage mit allen katalytisch aktiven Domänen zu interagieren[110]. Daher ist die Dynamik dieser Protein-Familie das Ziel vieler Untersuchungen durch die, die strukturellen Konformationsänderungen der PCP-Domäne, abhängig vom Interaktionspartner oder Beladungszustand, gezeigt werden konnten[111]. Die mittels [ $^{15}\text{N}$ ,  $^1\text{H}$ ]-TROSY (*transverse relaxation optimized spectroscopy*) untersuchte TycC<sub>3</sub>-PCP wechselt ihre Konformation abhängig von ihrem Beladungszustand. Die Apo-PCP liegt in der A-Form vor, während die Holo-PCP durch die H-Form beschrieben werden kann (Abb. 12). Beide Formen zeigen eine alternierende Änderung zu einer dritten Konformation, die für den beladenen und den unbeladenen Zustand identisch ist, die A/H-Form (Abb. 12). Die A-Form stellt die flexibelste der drei möglichen Formen da, mit langen ausgedehnten Loops. Die dritte Helix ( $\alpha_{\text{III}}$ ), die bei der A/H-Form als  $3_{10}$ -Helix vorliegt, formt in der A-Form einen Loop aus. Weitere Strukturelemente wie die Helix I, II und IV ( $\alpha_{\text{I}}$ ,  $\alpha_{\text{II}}$  und  $\alpha_{\text{IV}}$ ), liegen verkürzt vor, sind aber vergleichbar orientiert. Die Carbonylgruppe des konservierten Serinrestes (Tab. 2, T1) bildet eine Wasserstoffbrückenbindung zur Seitenkette des Arginins (R47) aus und fixiert so den Loop III. In der A/H-Konformation wird diese Wasserstoffbrücke zum Vorteil einer neuen zwischen der Seitenkette von Serin (S45) und Histidin (H44) aufgegeben, was den Loop III. Eine Serin zu Alanin Variation (S45A) arretiert die Apo-PCP in der A-Form[111]. Der Hauptunterschied zwischen der Holo-PCP spezifischen H-Form und der A/H-Form ist das Auflösen der dritten Helix ( $\alpha_{\text{III}}$ ). Gleichzeitig orientiert sich die Helix-IV parallel zu der Helix-I, was zur einer Neuorientierung der Helix-II, sowie des Loops II (zwischen  $\alpha_{\text{I}}$  und  $\alpha_{\text{II}}$ ) der am Serinrest (S45) die prosthetische Gruppe trägt, führt. Die Bewegung des Loops II während der Konformationsänderung zwischen der A/H- und der H-Form resultiert in einer weitreichenden Relokalisierung des PPant-Armes vom N-Terminus in der A/H- zum C-Terminus bei der H-Form um  $100^\circ$  und  $16 \text{ \AA}$ . Die Initialbeladung (*priming*) mit der prosthetischen Gruppe durch die Sfp erfolgt nach einer Interaktion der Apo-PCP in der A-Form mit der durch das Pseudodimer gebildeten aktiven Seite des Sfp (Abb. 12).

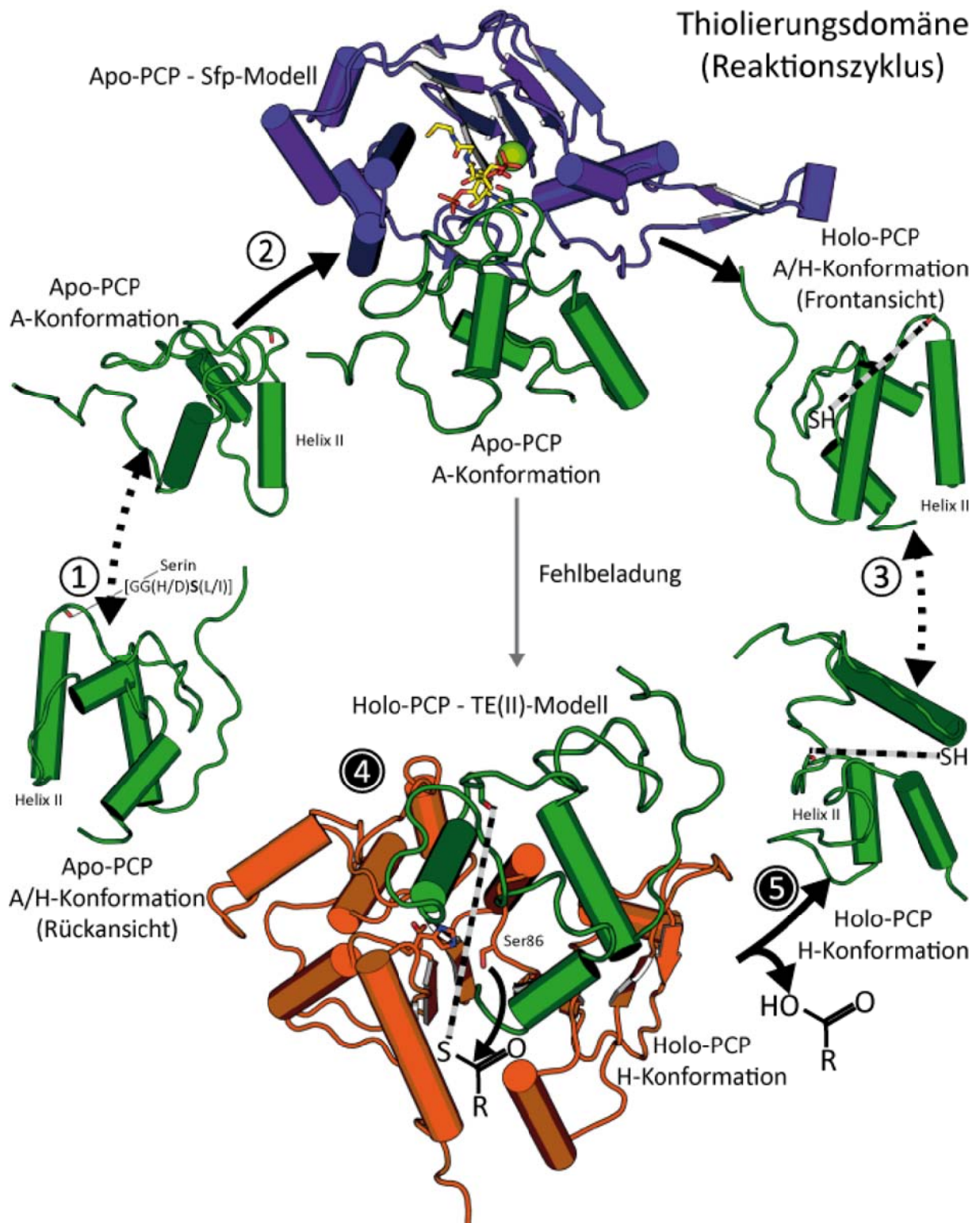


Abbildung 12: Reaktionszyklus der Thiolierungsdomäne (Peptidyl Carrier Protein – PCP) (1) Alternierende Konformationsänderung der Apo-PCP zwischen der A/H- (pdb: 2GDW) und der A-Form (pdb: 2GDY), der konservierte Serinrest ist als Stick-Modell dargestellt; (2) Priming-Interaktionsmodell zwischen der A-Form und Sfp (pdb: 2GE1); (3) Alternierende Konformationsänderung der Holo-PCP zwischen der A/H- (pdb: 2GDW) und der H-Form (pdb: 2GDX); **Bei Fehlbeladung:** (4) Interaktion der externen Thioesterase (Srf-TE II) mit der Holo-PCP in der H-Form (pdb: 2K2Q); (5) Abspaltung des falschen Substrates vom PPant-Arm der z.B. durch Acetyl-CoA fehlbeladenen Holo-PCP durch die katalytischen Triade (Asp189-His216-Ser86) der TE(II)-Domäne.

Da das CoA in der Zelle häufig als Acetyl- oder Malonyl-CoA vorliegt, kann die PCP mit einem acylierten PPant-Arm fehlbeladen werden. Der Clusterverbund vieler NRPS-Systeme kodiert daher für eine externe Thioesterase, die nachträglich durch die Interaktion mit der Holo-PCP in der H-Form den Thioester des PPant-Armes hydrolysiert (Abb. 12)[112]. Somit kann die Holo-PCP mit der freien Thiolgruppe durch die A-Domäne mit der korrespondierenden Aminosäure beladen werden und anschließend kann das Peptidylsubstrat durch die Kondensationsdomäne elongiert werden.

### iii.) Die Kondensationsdomäne (C-Domäne)

Die Kondensationsdomäne katalysiert die Elongationsreaktion durch den nukleophilen Angriff der  $\alpha$ -Aminogruppe der Aminosäure am Akzeptor-PCP (*downstream*) auf den Carbonyl-Kohlenstoff des Thioesters der Donor-PCP gebundenen Peptidkette. Neben der C-Domäne zeigen auch die Dualdomänen, die Kondensation und Heterozyklisierung katalysierende Zyklisierungsdomäne (Cy-Domäne)[113] als auch die Kondensation und Epimerisierung katalysierende C/E-Domäne[114] die gleiche Faltung. Allerdings weist das für die Katalyse zuständige, Kernmotiv (Tab. 3, C3) der Funktionalität entsprechend sequenzielle Abweichungen auf. Neben geringer sequenzieller und struktureller Unterschiede wird auch die Epimerisierungsdomäne durch diese Faltung beschrieben[115]. Innerhalb der C-Domänen sind weitere Untergruppen festzustellen. Abhängig von der, bei der Kondensation zwischen dem verwendeten Substrat der Donor-Seite ( $^D$ C) und der  $^L$ -Aminosäure der Akzeptor-Seite ( $C_A$ ), katalysierten Reaktion werden die C-Domänen in  $^L$ C $_L$ -Domänen ( $^L$ -Aminosäure-Peptidylrest),  $^D$ C $_L$ -Domänen ( $^D$ -Aminosäure-Peptidylrest) und Starter  $^S$ C $_L$ -Domänen ( $\beta$ -Hydroxycarbonsäure, meistens einer  $\beta$ -Fettsäure) unterschieden[116]. Die generelle Faltung der C-Domänen wird als ein Pseudodimer, bestehend aus jeweils einem N- und einem C-terminalen Bereich, beschrieben[117], die durch den, hinter dem Core-Motiv C4 (Tab. 3) liegenden, konservierten Serinrest getrennt werden.

Das Strukturmodell von VibH[117], einer alleinstehenden C-Domäne aus der Vibriobactin Biosynthese, zeigt, daß jede Domäne eine  $\alpha\beta\alpha$ -Sandwich-Faltung, die Homologien zu Chloramphenicol-Acetyltransferasen (CAT)[118-119] und E2-Einheit der Pyruvat-Acetyltransferasen[22] besitzt, annimmt. Die beiden Domänen sind V-förmig zueinander angeordnet, wobei zwei Regionen aus dem C-terminalen Bereich in den gegenüberliegenden N-terminalen



Bereich hineinragen. Die V-Form der beiden Bereiche hat ihre Spitze an dem konservierten Serinrest (Core-Motiv C4) und grenzt (strukturell) unmittelbar an den *Floor Loop* (Core-Motiv C5). Zwischen dem *Floor Loop* und der ebenfalls zwischen den Bereichen liegenden *Bridge-Region* verläuft der Reaktionskanal (*solvent channel*) in dessen Mitte sich das konservierte Histidin (Core-Motiv C3) befindet und an dessen Enden der jeweilige Donor bzw. Akzeptor-Eingang liegt (Abb. 13).

Core-Motiv	Pos.	Sequenz	Funktion
<b>C-Dom.</b>	C1	<u>Gln</u> 24 Sx <b>AQ</b> xR(LM)(WY)xL	Strukturelement (Acceptor Seite)
	C2	<u>Arg</u> 68 RHExLRTxF	Strukturelement
	C3	<u>His</u> 142 HHxxxDG	Katalyseelement
	C4	<u>Ser</u> 176 YxD(FY)AVWx <sub>6</sub> <b>S</b>	Strukturelement ( <i>N- und C-Term. Grenze</i> )
	C5	<u>Gly</u> 288 (IV) <b>G</b> xFVNT(QL)(CA)xR	Strukturelement ( <i>floor loop</i> )
	C6	<u>Pro</u> 344 (HN)QD(YD) <b>PFE</b>	Strukturelement (cross over)
	C7	<u>Arg</u> 338 RDxSRNPL	Strukturelement (Acceptor Seite)

Tabelle 3: Die Core-Motive der Kondensationsdomänen sowie ihre Verteilung und Funktion.

Bevor die Kondensationsreaktion stattfinden kann, muß die Akzeptor-PCP durch die im gleichen Modul enthaltene A-Domäne (bei NRPS Typ A) mit der korrespondierenden Aminosäure beladen werden. Die so beladene Akzeptor-PCP interagiert intramodular mit der C-Domäne[120]. Anschließend wechselwirkt die mit dem Amino-Peptidylrest beladene Donor-PCP intermodular mit der C-Domäne, wobei der Peptidylrest auf die  $\alpha$ -Aminosäure der Akzeptor-PCP übertragen und dadurch um eine Aminosäure elongiert wird.

Die Interaktionen zwischen der C- und PCP-Domäne konnten strukturell bislang nicht gezeigt werden, da die bidomänale TycC<sub>5-6</sub> PCP-C Struktur (Abb. 13) nur eine unproduktive Interaktion vor der eigentlichen Kondensation zeigt[121]. Dabei ist das konservierte Serin der PCP ca. 49 Å vom konservierten, katalytisch aktiven Histidin der C-Domäne entfernt und kann somit nicht durch die volle Reichweite (20 Å) des PPant-Armes abgedeckt werden. Sobald die Kondensationsreaktion durchgeführt wurde wird die Donor-PCP durch die eigene A-Domäne erneut acyliert. Der elongierte Peptidylrest der Akzeptor-PCP kann dann entweder durch eine zusätzliche Domäne modifiziert, durch eine Kondensationsdomäne elongiert oder an die Thioesterase, die eine hydrolytische bzw. zyklische Abspaltung des Endproduktes katalysiert,

weitergereicht werden.

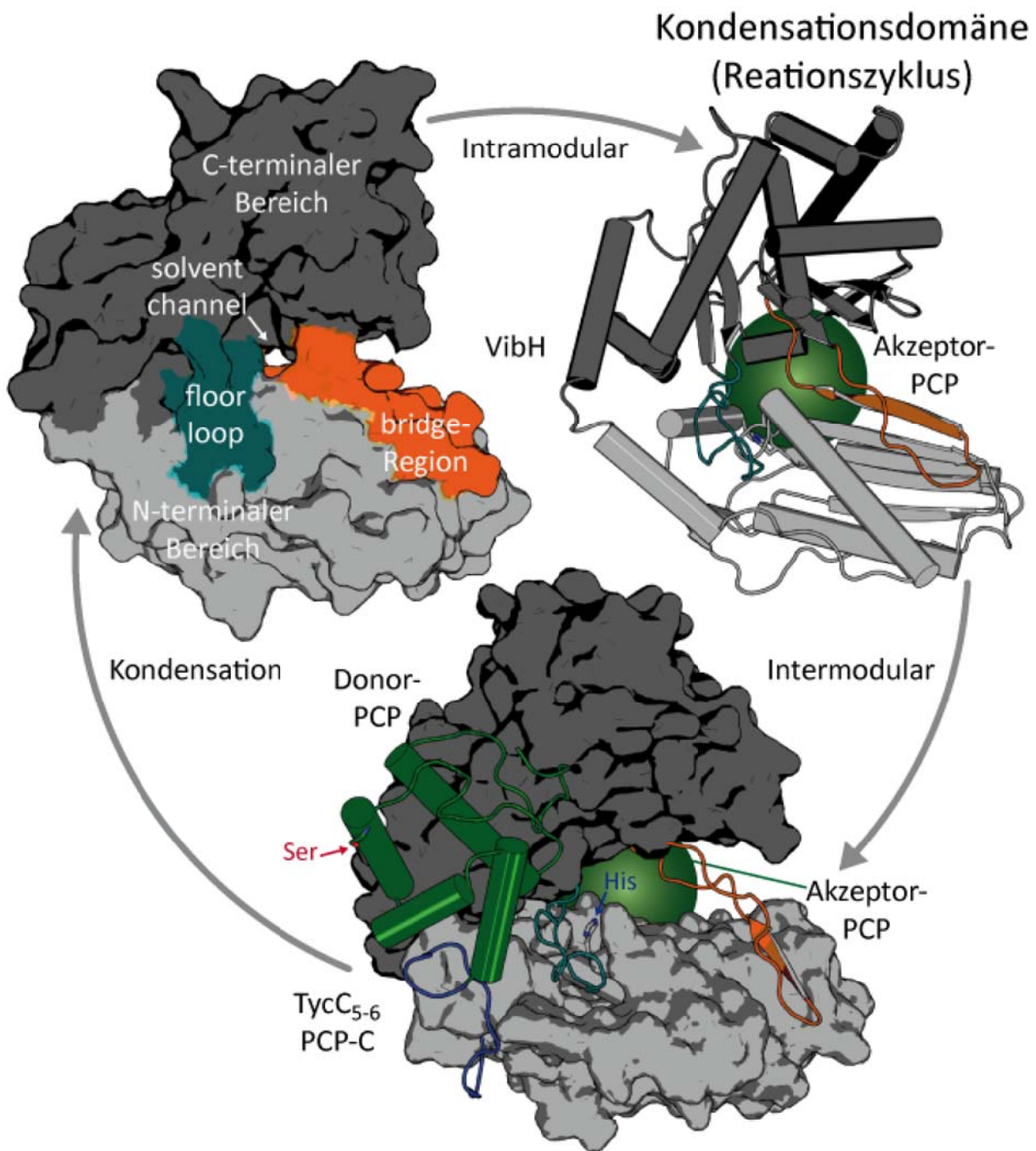


Abbildung 13: Reaktionszyklus der Kondensations-Domäne (C-Domäne) am Beispiel von VibH (pdb:1L5A) und TycC<sub>5-6</sub> PCP-C (pdb:2JGP).

## iv.) Die Thioesterase (TE-Domäne)

Die ca. 250 Aminosäuren lange, terminale Thioesterase ist im Gegensatz zu den anderen essentiellen Domänen nicht in die einzelnen Module der NRPS-Systeme integriert, sondern



ausschließlich in dem terminalen Modul als letzte Domäne der nichtribosomalen Peptid-synthetasen vorhanden. Durch ihre Position ist es ihr möglich, die Freisetzung der elongierten Peptidkette zu katalysieren, damit diese ihre biologische Aktivität entfalten können. Die Produktfreisetzung kann dabei mechanistisch[122] auf mannigfaltige Weise erfolgen (Abb. 9-1). Strukturell gehören die Thioesterasen zu der Familie der  $\alpha/\beta$ -Hydrolasen[123] und katalysieren neben der hydrolytischen Abspaltung auch die regio- und stereospezifische Freisetzung als Makrolakton- bzw. Makrolaktam-Ring. Während die typischen Vertreter der Familie der  $\alpha/\beta$ -Hydrolasen durch eine charakteristische Faltung (abwechselnd  $\beta$ -Faltblatt und  $\alpha$ -Helices) beschrieben werden können, ist bei den Thioesterasen die  $\alpha$ -Helix D durch eine dynamische *Lid*-Region ersetzt (Abb. 14). Die enzymatische Wirksamkeit wird durch die katalytische Triade aus Serin (Nukleophil), Histidin (Wasserstoff-Akzeptor) und Aspartat (Carboxylat) ermöglicht. Der konservierte Serinrest (Tab. 4) der Loop-Region wird durch das  $\beta$ -Faltblatt 5 ( $\beta 5$ ) und die  $\alpha$ -Helix C ( $\alpha C$ ) begrenzt und ist zwischen dem ca. 19 Å langen Substratkanal (*substrate channel*), der zum konservierten Serin der PCP-Domäne (Abb. 14, Interaktion) führt, und der hydrophoben Vertiefung lokalisiert[124].

Core-Motiv	Pos.	Sequenz	Funktion
<b>TE</b>	TE1	<u>Ser80</u>	GxSxG Katalyseelement: Oxoester Intermediat

Tabelle 4: Das konsevierte Motiv der Thioesterase-Domäne (TE).

Die konservierten Aminosäuren Serin ( $\beta 5$ - $\alpha C$ ), Histidin ( $\beta 8$ - $\alpha F$ ) und Aspartat ( $\beta 6$ -*Lid*-Region) sind in der offenen Konformation der *Lid*-Region (Abb. 14, SrfAC-TE Offen) Wassermolekülen zugänglich. Das als Thioester am PPant-Arm der PCP-Domäne gebundene Peptidyl-Substrat wird auf das, durch die katalytische Triade deprotonierte, Serin (Oxyanion) übertragen. Während der PCP-TE-Interaktion ist die *Lid*-Region über dem Oxyanion positioniert (Abb. 14, Interaktion). Untersuchungen des bidomänen Proteins EntF PCP-TE zeigen einen sehr langsamen NH-Austausch der aus  $\beta$ -Faltblättern bestehenden Kernstruktur und den umgebenden  $\alpha$ -Helices[125]. Dagegen zeigt die *Lid*-Region einen regen NH-Austausch, was die Beweglichkeit dieser Region während der Interaktion bestätigt. Durch die Interaktion mit der PCP wird das als Thioester gebundenes Peptidyl-Substrat auf das Oxyanion übertragen. In der geschlossenen Konformation (Abb. 14, SrfAC-TE Geschlossen) können Wassermoleküle

nur in einen kleinen Bereich der hydrophoben Vertiefung eindringen[124]. Somit wird das nun als Oxoester vorliegendes Peptidyl-Substrat vor der spontane Hydrolyse geschützt (Abb. 14). Im Gegensatz zu der offenen Konformation zeigt die zweite  $\alpha$ -Helix eine große Bewegung zum aktiven Zentrum hin, was in der Ausbildung der geschützten hydrophoben Vertiefung resultiert.

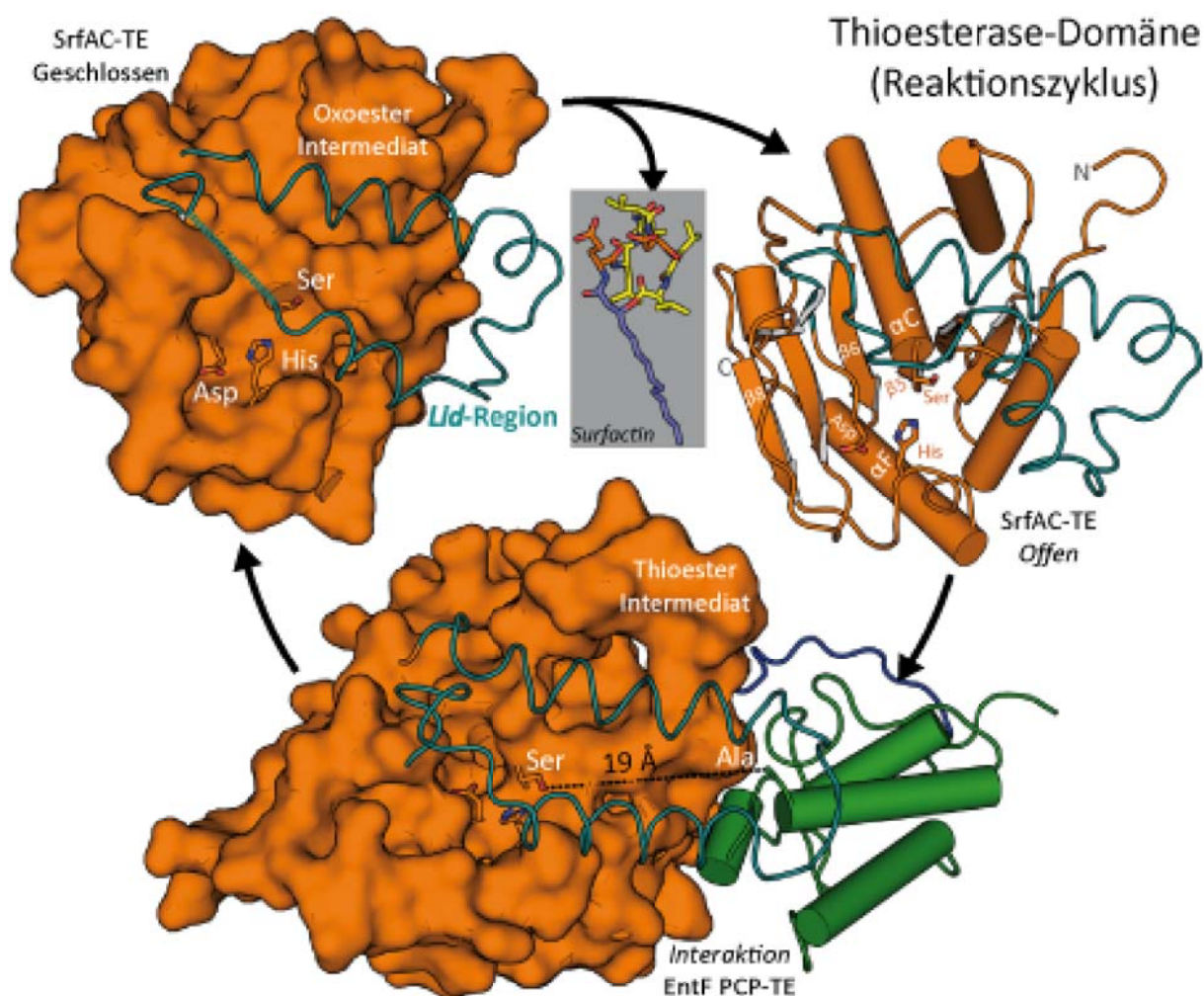


Abbildung 14: Reaktionszyklus der Thioesterase-Domäne (TE-Domäne) am Beispiel der geschlossenen und offenen Form der Surfactin (pdb:2NPV) produzierenden SrfAC-TE (pdb:1JMK) und der bidomänen EntF PCP-TE (pdb:2ROQ).

Strukturmodelle der PKS-TE aus der Pikromycin Biosynthese zeigen die Thioesterase sowohl mit dem kovalent gebundenen phosphorylierten Substratanalogon (pdb: 2H7Y) als auch mit dem reduzierten Substrat (pdb: 2H7X)[126]. Dabei zeigt das Substratanalogon eine Komplexierung in der hydrophoben Vertiefung, wobei das Oxyanion geschützt unter der *Lid*-

Region liegt. An fast allen polaren Interaktionen innerhalb der hydrophoben Vertiefung sind durch am Peptidrückgrat oder den wenigen polaren Seitenketten fixierte Wassermoleküle beteiligt. Dass nur wenige direkte Interaktionen zwischen dem Substrat und den umliegenden Aminosäuren zu verzeichnen sind, erklärt die Promiskuität dieser TE-Domäne[126]. Am Beispiel des reduzierten Substrates zeigt sich die Notwendigkeit der erneuten Konformationsänderung der *Lid*-Region um das zyklisierte Substrat freizusetzen. Für NRPS-Systeme konnte bislang die Interaktion zwischen dem Substrat und der entsprechenden TE-Domäne noch nicht gezeigt werden. Allerdings muss das als Oxoester gebundene Substrat einen internen nukleophilen Angriff ausführen, um den Ringschluss zu bilden. Die Surfactin-produzierende Thioesterase formt einen Makrolaktonring zwischen der freien Carboxylgruppe des letzten Leucinrestes und der  $\beta$ -Hydroxygruppe der Fettsäure (Abb. 14, grauer Kasten, Fettsäure: blau, hydrophile Aminosäuren: gelb, saure Aminosäuren: orange). Durch den anschließenden Wechsel in die offene Konformation kann das fertige Produkt der nichtribosomalen Synthetase freigesetzt werden. Bei iterativen Thioesterasen wird durch den *Backward*-Mechanismus[127] der als Oxoester gebundene Peptid-Monomer zuerst mit der auf der PCP lokalisierten identischen Peptid-Monomer ligiert und anschließend erneut auf die Thioesterase übertragen. Anschließend wird der Peptidoligomer dann, ähnlich wie bei linearen TE-Domänen, durch „*Head to tail*“[127]- oder Seitenketten-Zyklisation[94] abgespalten.

## v.) *Modifizierende 'in cis'-Domänen (Substitution)*

Die Heterozyklisierung- und Kondensation-katalysierenden Tandem-Cy-Domäne zieht große Sequenzhomologien zu C-Domänen[116], obwohl sich ihre Core-Motive stark von den Core-Motiven der C-Domänen unterscheiden (Tab. 5). Diese Domäne katalysiert neben der Kondensationsreaktion die Bildung eines Heterozyklus zwischen Serin, Threonin oder Cystein an der Donorseite und einer beliebigen Aminosäure an der Akzeptorseite, resultierend in einem Oxazolin-, Methyloxazolin- bzw. Thiazolinring (Abb. 9-2)[128]. Diese können durch eine z.B. in die A-Domäne integrierte Ox-Domäne weiter oxidiert werden[87]. Die C-Domänen können außerdem durch die strukturell verwandten Tandem-C/E-Domänen substituiert werden, die durch ein zweites Katalyseelement von den konventionellen C- und E-Domänen unterschieden werden können[114].

# STRUKTURMODELL DES TERMINATIONSMODULS DER SURFACTIN SYNTHETASE

Core-Motiv	Pos.	Sequenz	Funktion
<b>Cy-Dom.</b>	Cy1 <u>Phe</u> 23	FPL(TS)xxQxAYxxGR	
	Cy2 <u>Arg</u> 73	RHx(IM)L(PAL)x(ND)GxQ	
	C3 <u>Asp</u> 150	D(NLI)xDxxS	Katalyseelement
	Cy3 <u>Pro</u> 216	LPxxPxLPLxxxP	
	Cy4 <u>Trp</u> 277	(TS)(PA)x <sub>3</sub> (LAF)x <sub>6</sub> (IVT)LxxW	
	Cy5 <u>Leu</u> 312	(GA)DFTxLxLL	
	Cy6 <u>Pro</u> 365	PVVFTSxL	
	Cy7 <u>Trp</u> 400	(ST)(QR)TPQVx(LI)Dx <sub>13</sub> WD	
<b>R-Dom.</b>	R1 <u>Tyr</u> 110	V(LL)TGATG(FL)GxxLL	
	R2 <u>Arg</u> 156	VxLLVRA	
	R3 <u>Gly</u> 192	GPLGxPxLGL	Glycine-rich Motiv (PPi-Bindung)
	R4 <u>Tyr</u> 221	VYPYxYLxxPNVxxT	Putative Nukleotid-Bindung
	R5 <u>Lys</u> 290	GYxxSKWAAE	
	R6 <u>Arg</u> 309	RPG	
	R7 <u>Tyr</u> 496	Yx <sub>4</sub> G(LF)LxxP	

Tabelle 5: Die Core-Motive der in cis modifizierenden Domänen (Substitution).

Die Freisetzung der Peptidyl-Vorläufer kann auf vielfältig Weise erfolgen, was zu der hohen strukturellen Diversität von NRPS-Produkten beiträgt. Die Reduktase-Domäne kann die Thioesterase substituieren und die Peptidyl-Vorläufer nicht als Makrozyklen sondern als C-terminale Aldehyde oder sogar Alkohole freisetzen[84]. Die Zweielektronen-Reduktase ( $R^{2e-}$ -Domäne) reduziert das Peptidyl-Substrat zunächst zum Aldehyd der anschließend von einer zweiten Reduktase wie bei linearem Gramicidin zum Alkohol reduziert werden kann. Die Vierelektronen-Reduktase ( $R^{4e-}$ -Domäne) reduziert das Substrat über den Aldehyd direkt zum Alkohol wie beim Siderophor Myxochelin[129]. Die Reduktase-Domänen sind NAD(P)H-abhängig und weisen starke Sequenzhomologien zu der *Short-Chain* Dehydrogenase/Reduktase Superfamilie (SDR)[92]. Diese Domänen zeigen die für Nukleotid bindende Proteine charakterische Rossmann-Faltung[130] mit dem entsprechenden Nukleotid-bindenden sowie Phosphatrest-bindenden Motiv (Tab. 5, R4, R3). Die Rossmann-Faltung besteht aus einem zentral angeordnetem sechs bis sieben strängigem  $\beta$ -Faltblatt mit einer 3-2-1-4-5-6-7 Topologie und einem langem *Crossover*-Bereich zwischen Strang drei und vier.

Weitere drei bis vier  $\alpha$ -Helices auf jeder Seite schirmen das zentrale  $\beta$ -Faltblatt-Motiv ab. Neben der reduktiven Freisetzung linearer Peptidyl-Produkte, wurde in PKS/NRPS-Hybridsystemen, wie bei Tenellin[131] auch Reduktase-katalysierte Dieckmann-Kondensationen zur Freisetzung der Peptidyl-Produkte beschrieben. Hierbei handelt es sich um nichtreduktive Abspaltungen zyklischer Produkte.

## vi.) Modifizierende 'in cis'-Domänen (Insertion)

Neben den C/E-, Cy- und R-Domänen, die essentielle C- bzw. TE-Domänen substituieren, können NRPS-Systeme zusätzliche, ergänzende Domänen erhalten, die die Mannigfaltigkeit der Peptidyl-Produkte zusätzlich erweitern (Abb. 9-3). Dabei sind die Ox- sowie die  $N$ -Mt-Domänen in die A-Domäne zwischen den A8 und A9 Kernmotiven inseriert (Tab. 6).

Core-Motiv	Pos.	Sequenz	Funktion
<b>Ox-Dom.</b>	A8 <u>Asp</u> 389	GRxDxQVKIRGxRIELGEIE	
	Ox1 <u>Lys</u> 100	$x_{140}$ -KYxYxSxGxxY(PG)VQ	
	Ox2 <u>Gly</u> 121	GxxxG(LV)xxGxYYY/HD)P	
	Ox3 <u>Tyr</u> 157	IxxxYG- $x_{95}$	
	A9 <u>Pro</u> 764	LPxYM(IV)P	
<b><math>N</math>Mt-Dom.</b>	A8 <u>Asp</u> 389	GRxDxQVKIRGxRIELGEIE	
	Mt1 <u>Gly</u> 86	$x_{110}$ -VL(DE)xGxGxG	
	Mt2 <u>Ser</u> 265	NELsYRYxAV	
	Mt3 <u>Arg</u> 395	VExSxARQxGxLD- $x_{55}$	
	A9 <u>Pro</u> 846	LPxYM(IV)P	
<b>E-Dom.</b>	E1 <u>Pro</u> 12	PIQxWF	
	E2 <u>His</u> 137	HHxISDG(WV)S	Katalyseelement
	E3 <u>Asp</u> 249	DxLLxAxG	
	E4 <u>Glu</u> 274	EGHGRE	Strukturelement ( <i>Floor Loop Start</i> )
	E5 <u>Thr</u> 311	RTVGWFTxxYP(YV)PFE	Strukturelement ( <i>Floor Loop Ende</i> )
	E6 <u>Pro</u> 340	PxxGxGYG	
	E7 <u>Phe</u> 368	FNYLG(QR)	

Tabelle 6: Die Core-Motive der *in cis* modifizierender Domänen (Insertion)[78].

Die in Zusammenhang mit den Cy-Domänen häufig auftretenden Ox-Domänen katalysieren

in einer FMN-abhängigen Reaktion die Oxidation der Oxazolin- oder Thiazolin-Zwischenprodukte[87]. Die *N*-Methylierung ist neben der Makro- und Heterozyklisierung eine weitere Strategie, um die Protease-Affinität zu senken und damit eine höhere Beständigkeit und Wirksamkeit zu erreichen. Bei dieser modifizierenden Reaktion wird die Methylgruppe vom Kosubstrat SAM über einen Radikal-Mechanismus auf eine NH-Gruppe übertragen[132]. Neben diesen bislang strukturell nicht aufgeklärten modifizierenden Domänen ist die Epimerisierungs-Domäne die in NRPS-Systemen am häufigsten vorkommende Modifikations-domäne. Zusätzlich zu der katalytischen Funktion ist der C-terminale Teil dieser Domäne an der Wechselwirkung zweier nichtribosomaler Peptidsynthetasen beteiligt[133]. Die Epimerisierungsdomäne (E-Domäne) wird daher häufig am C-terminalem Ende einer Synthetase kodiert, wobei die Wechselwirkung zu der folgenden Synthetase durch die communication-mediating Domäne (COM-Domäne) vermittelt wird. Analog zu der Struktur der Kondensationsdomäne zeigt das Strukturmodell der Epimerisierungsdomäne ebenfalls einen *Floor Loop* sowie die *Bridge-Region* (Abb. 15; pdb: 2XHG).

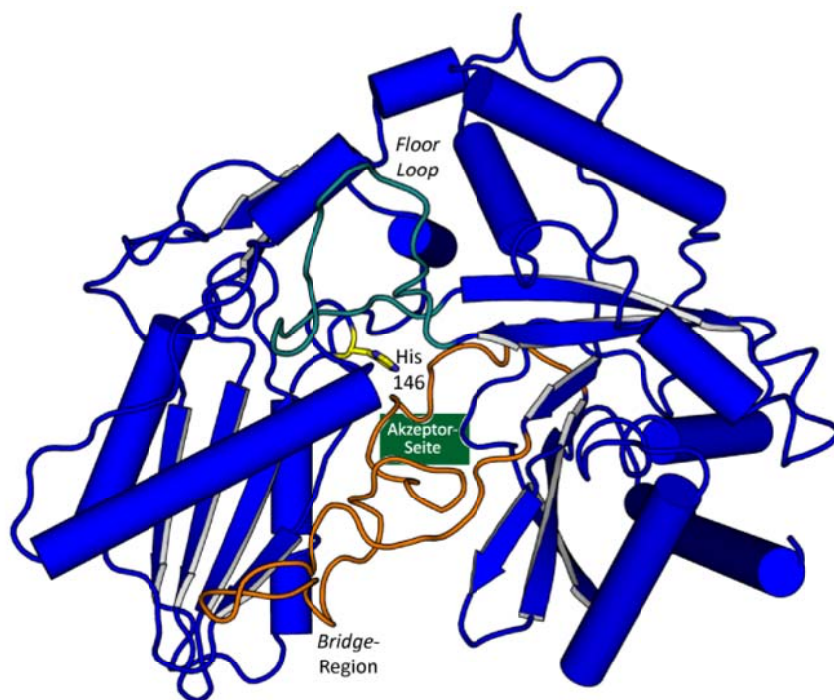


Abbildung 15: Strukturmodell der TycA Epimerisierungs-Domäne (pdb: 2XHG).

Beim strukturellem Vergleich mit der Kondensationsdomäne TycC<sub>6</sub>-C (pdb: 2JGP) zeigt der *Floor Loop* eine Ausdehnung in Richtung der Donor-Seite der TycA E-Domäne. Ob dieser Bereich an der Interaktion mit der PCP-Domäne beteiligt ist, wurde bislang noch nicht

geklärt. Bei Sequenz-Untersuchungen wurde eine Abweichung in der Sequenz des PCP-Core-Motivs in Modulen mit einer E-Domäne festgestellt (Tab. 2; PCP<sup>C</sup>: GGHSL, PCP<sup>E</sup>: GGDSI)[134], die in der Interaktionsfläche der PCP-Domäne liegt[125]. Die bereits durch Sequenzvergleiche postulierte Erweiterung der *Bridge*-Region durch die Insertion von 12 Aminosäuren ist nicht in das  $\beta$ -Faltblatt-Motiv des N-terminalen Bereichs eingebunden, wie bei den C-Domänen bereits gezeigt wurde[117], weist aber Interaktionen mit dem  $\beta$ -Strang  $\beta 2$  auf. Während die *Bridge*-Region bei den C-Domänen für die Ausformung des Reaktionskanals (*solvent channel*) verantwortlich ist (Abb. 13), blockiert die erweiterte *Bridge*-Region in der E-Domäne den Zugang von der Akzeptor-Seite und formt somit eine, nur einseitig zugängliche, Reaktionskammer mit dem katalytisch relevanten Histidin<sup>146</sup> (Abb. 15) aus.

## vii.) Modifizierende 'in trans'-Domänen

Die große Vielfalt der NRPS-Produkte wird durch modifizierende Enzyme die nicht in der gleichen Polypeptidkette kodiert werden, sondern *in trans* wirken, erhöht. Die verschiedenen Modifikationen können nach, während[88], oder, wie z.B. bei der Asparagin Oxygenase AsnO[135], vor der Peptidsynthese durch ein NRPS-System erfolgen. Die Asparagin-Oxygenase AsnO gehört zu der CSL Superfamile[136] (*clavaminate synthase like superfamily*) und zeigt hohe strukturelle Homologien zu CAS (*clavulanic acid synthase*)[137]. Die Kernstruktur besteht aus einer neun  $\beta$ -strängigen ( $\beta A$ - $\beta I$ ) *Jelly Roll*-Topologie, bestehend aus einem Haupt- und einem Neben- $\beta$ -Faltblatt mit fünf ( $\beta B$ ,  $\beta G$ ,  $\beta D$ ,  $\beta I$  und  $\beta C$ ) bzw. drei ( $\beta F$ ,  $\beta E$  und  $\beta H$ ) Strängen. Das *Jelly Roll*-Motiv ist durch zwei lange  $\alpha$ -helicale Unterdomänen eingebunden, die N-terminale Region (S10-G71,  $\alpha 1$ - $\alpha 3$ ) und dem Insertionsbereich (V188-P274,  $\alpha 5$ - $\alpha 7$ ). Der mit der Apo-Form beinahe identische AsnO-Fe<sup>2+</sup>-Komplex zeigt im aktiven Zentrum ein oktaedrisch koordiniertes Fe<sup>2+</sup>-Ion und eine offene *Lid*-Region (Abb. 16, rot), die den Zugang zum aktiven Zentrum ermöglicht. In einem Produkt-Komplex sind im aktiven Zentrum das (2S,3S)-3-Hydroxyasparagin (hAsn) sowie das nach der Hydroxylierung von Asparagin aus  $\alpha$ -Ketoglutarat entstandene Succinat komplexiert (Abb. 16). Das aktive Zentrum ist nicht mehr zugänglich, was durch die Konformationsänderung der *Lid*-Region zu erklären ist (Abb. 16, blau). Das durch die AsnO katalysierte Reaktion entstandene 3-Hydroxyasparagin wird an Position 9 vom 11-gliedrigen Lipopeptid-Reinges des Makrolakton-Antibiotikums CDA (*calcium-dependent antibiotic*) eingebaut.



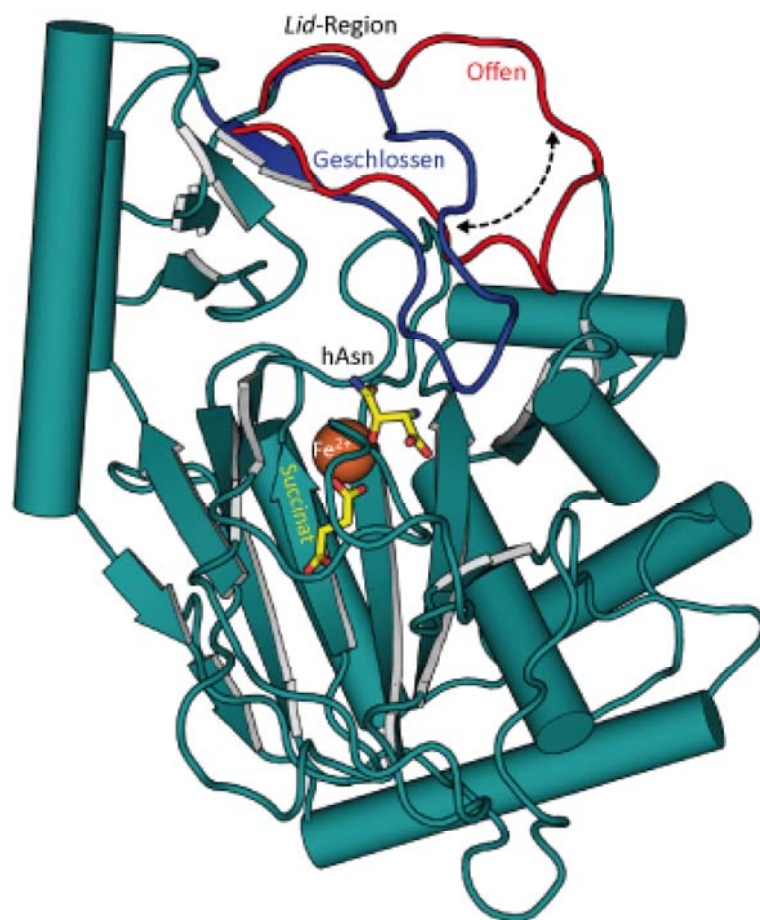


Abbildung 16: Strukturmodell der Asparagin-Oxygenase AsnO mit der flexiblen Lid-Region. Komplex mit Fe(II), (2S,3S)-3-Hydroxyasparagin (hAsn) und Succinat (pdb: 2OG7). Die flexible Lid-Region ist blau dargestellt und zum Vergleich ist die Lid-Region der Apo-Form rot gezeigt (pdb: 2OG6).

## 2 Aufgabenstellung

Die Intention dieser Arbeit war es durch verschiedene Methoden der makromolekularen Strukturaufklärung, wie z.B. Proteinkristallisation, Kleinwinkelstreuung sowie CryoEM, Daten für die vollständige oder partielle Aufklärung der globalen Architektur der nichtribosomalen Peptidsynthetasen zu erhalten. Zusammen mit den bereits vorhandenen Daten aus der Strukturaufklärung einzelner Domänen durch Proteinkristallisation oder NMR-Experimente sollte ein Bild über die räumliche Verteilung sowohl von einzelnen Domänen innerhalb eines Modules als auch der Anordnung der Module zueinander bestimmt werden (Abb. 17).

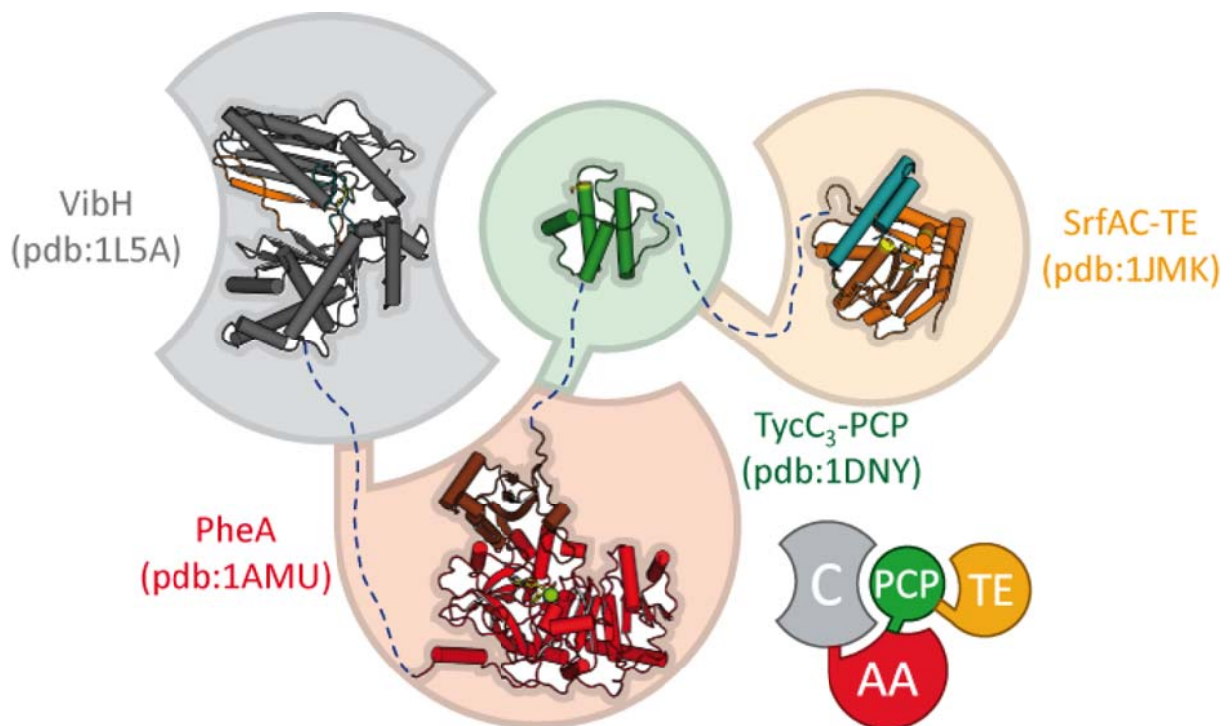


Abbildung 17: Representative Darstellung eines terminalen NRPS-Moduls.

Viele Untersuchungen im Vorfeld dieser Arbeit haben gezeigt, dass mit der Funktion der einzelnen Domänen, wie der Dynamik der PCP- und der Sub-A-Domäne sowie der flexiblen *Lid*-Region der TE-Domäne oder der *Bridge*-Region der C-Domäne, ein hohes Maß an struktureller Variabilität einhergeht. Neben der strukturellen Vielfalt der einzelner Bereiche der NRPS-Domänen ist bislang auch sehr wenig über die Linkerbereiche zwischen den einzelnen funktionalen Domänen der NRPS-Systeme bekannt. Untersuchungen an *Swinging*-Domänen enthaltenden multimodularen Enzymen und Komplexen, wie an dem PDH-E2-

Komplex oder den FAS- und PKS-Systemen zeigten eine hohe Population der Prolin- und Alaninreste in den Linkerbereichen der *Swinging*-Domänen. Die Linkerbereiche der NRPS-Systeme konnten aus Mangel an globalen Struktur-Informationen bislang nicht systematisch untersucht werden. Des weiteren ermöglicht bereits jetzt die Kombination aus chemischen und biologischen Techniken, durch das Einsetzen synthetisch sowie chemoenzymatisch hergestellter Intermediate den Zugang zu neuen Metaboliten[62]. Die Erkenntnisse über die globale Architektur der NRPS-Systeme steigern nicht nur das Potential solcher Ansätze, sondern erweitern die Produktpalette dieser Systeme durch erleichterte Erschließung neuer und synthetischer Metaboliten[138].

Durch die strukturellen Untersuchungen in dieser Arbeit soll des weiteren geklärt werden, ob die aktiven Zentren innerhalb eines NRPS-Modules in der Reichweite des verknüpften PPant-Armes (20 Å) liegen und die Modul-interne Produkt-Synthese dadurch ohne globale Konformationsänderungen auskommt, oder die globalen Konformationsänderungen, vergleichbar mit den FAS- und PKS-Systemen, für die Produkt-Synthese essentiell sind[39]. Zusätzlich zeigen Experimente an Carrier-Proteinen signifikante Änderungen in ihrer Konformation, abhängig von ihrem Beladungszustand[139-140]. Die Arretierung der vorher dynamischen TycC<sub>3</sub>-PCP-Domäne in der A-Konformation (Apo-Konformation) kann durch die Variation des konservierten Serinsrestes zu Alanin erfolgen[111]. Daher sollte dieser Ansatz als Ausgangspunkt für weiterführende Experimente gewählt werden. Darüber hinaus soll die bereits untersuchte PCP-Arretierung durch verschiedene Derivate des nicht hydrolisierbaren Amino-PPant-Armes erreicht werden.

### 3 Material

#### 3.1 Equipment

Bezeichnung ( <i>Hersteller</i> )	Gerät
Autoclavi SPA FVA3/A1 ( <i>Fedegari</i> )	Autoklav
Sartorius	Analysewaagen
DESY – Hamburg (Deutschland)	Beamline X13 , X33 und BW7
ESRF – Grenoble (Frankreich)	Beamline 14-2/14-3/14-4
SLS – Villingen (Schweiz)	Beamline X06SA
Seralpur Pro90CN ( <i>Seral</i> )	Bidestillationsanlage
Inkubator 2736 ( <i>Köttermann</i> )	Brutschrank
BSK ( <i>Antair</i> )	Clean-Bench
CAMEDIA C-3030 Zoom ( <i>Olympus</i> )	Digitalkamera
Videokamera CS1 ( <i>Cybertech</i> )	Dokumentation von
Thermodrucker Video Copy Processor ( <i>Mitsubishi</i> )	DNA-Agarosegelen
Crystal Pro HT 110 ( <i>TriTek</i> )	Dokumentationssystem
Agarose-Gelkammer ( <i>Feinmechanik, Philipps-Uni MR</i> )	Elektrophoresekammern
Gene Pulser und Pulse-Controller ( <i>Bio-Rad</i> )	Elektroporations-Pulser
Agilent series 1100 ( <i>Hewlett Packard</i> )	ESI-MS-System
EmulsiFlex-C5 ( <i>AVESTIN</i> )	Fluidizer
HiLoad Superdex Gelfiltrationssäule	FPLC-Säulen
Ni-NTA Säulen, HiTrap Desalting ( <i>GE Healthcare</i> )	
FPLC-System 250 ( <i>Pharmacia/GE-Healthcare</i> ): Gradienten-Programmer GP-250, Pumpe P-500, Uvicord Opt. Einheit UV-1 (l = 280 nm), Uvicord Kontrolleinheit UV-1, 2-Kanal Flachbett-schreiber REC-102, Injektionsventil V-7, 3-Wege-Magnetventil PSV-100, Fraktionssammler FRAC-100	FPLC-System
French-Pressure Cell-Version 5.1 ( <i>Sim Aminco</i> ), 20k Rapid-fill cell (40 mL)	French Press
Thermomixer Comfort ( <i>Eppendorf</i> )	Heizschüttler
Agilent series 1100 HPLC-System ( <i>Hewlett Packard</i> ): Vakuum-Entgaser, DAD-Detektor, Quaternäre Pumpe, Autosampler, Fraktionskollektor, HP-Chemstation	HPLC-System

CC 250/2 Nucleodur 100-3 C18ec	HPLC-Trennsäulen
CC 250/3 Nucleosil 120-3 C18 ( <i>Marcherey &amp; Nagel</i> )	
Cartesian Microsys SQ 4000 ( <i>Genomic Solutions</i> )	Kristallisationsroboter
Mikroskop: SZ60, Lichtquelle: TH3, Objektivaufsatz: SZ-PT ( <i>Olympus</i> )	Kristallobservation
Series 25 Incubator Shaker ( <i>New Brunswick Scientific</i> )	Luftschüttler
FLEX III ( <i>Bruker</i> )	MALDI-TOF
EmulsiFlex-C5 ( <i>Avestin</i> )	Microfluidizer
ND-1000 ( <i>PEQLab</i> )	Nanodrop
Mastercycler Personal ( <i>Eppendorf</i> )	PCR-Maschine
APEX396 Synthesizer ( <i>Advanced ChemTech</i> )	Peptid-Synthetisierer
CG 840 ( <i>Schott</i> )	pH-Meter
Ultraspec 3000 ( <i>Pharmacia Biotech</i> )	Photometer
Lissy ( <i>Zinsser Analytic</i> )	Pipettierroboter
Research Pipettes: Labmate L5000 ( <i>Abimed</i> ), 2.5 µL, 10 µL, 20 µL, 100 µL, 200 µL, 1 mL ( <i>Eppendorf</i> )	Pipetten
Äkta Prime ( <i>Amersham Pharmacia Biotech</i> )	Proteinaufreinigung
2-Kanal Flachbett-schreiber REC-112	Schreiber
Mini-Protean 3 ( <i>Bio-Rad</i> )	SDS-PAGE-Gelkammer
Speed Vac ( <i>Savant</i> ), Univapo 150 ( <i>Uniequip</i> )	Speed-Vac
Thermal Cycler 480, Gene Amp PCR System 2400, Gene Amp PCR System 9700 ( <i>Perkin-Elmer</i> )	Thermocycler
Vortex Genie2 ( <i>Scientific Industries</i> )	Vortexer
Aquatron Rundsüttelmaschine ( <i>Infors</i> )	Wasserbad
Biofuge pico ( <i>Heraeus</i> ), RC 26 plus - Rotoren SS34 und SLA300, RC 5B Plus ( <i>Sorvall</i> ), Megafuge 1.0R, Minifuge RF, Centrifuge 5415D Centrifuge 5810R - Rotor A-4-62 ( <i>Eppendorf</i> )	Zentrifugen

Tabelle 7: Verwendetes Equipment

### 3.2 Chemikalien, Enzyme und Verbrauchsmaterial

Hersteller (Vertriebsort)	Produkte
Amersham Biosciences European GmbH (Freiburg); GE Healthcare	Ampicillin, IPTG, Kanamycin, Hefeextrakt, Coomassie Brilliant Blue G R250, Agar Nr.1, HiTrap™-Entsalzungssäulen

<i>Bachem</i> (Weil am Rhein)	geschützte und ungeschützte Aminosäuren
<i>Biomol</i> (Ilvesheim)	DTT
<i>Brand</i>	Einmalküvetten Plastbrand PS
<i>Eppendorf</i> (Hamburg)	1.5 mL und 2.0 mL Reaktionsgefäße
<i>Eurogentech</i> (Seraing, Belgien)	Agarose
<i>Fermentas</i> (St. Leon-Rot)	PageRuler Unstained Protein Ladder
<i>Finnzymes</i> (Espoo, Finnland)	Phusion DNA-Polymerase, DyNAzyme
<i>Fluka</i> (Neu Ulm)	SDS, PEG 4000, PEG 8000, HEPES, Tris, MES, Na-Acetat, Na-Cacodylat, Na-Citrat, Mg-Acetat, Mg-Chlorid, Glycerol, Isopropanol, Ammoniumphosphat
<i>GATC</i> (Konstanz)	Plasmidsequenzierungen
Hempton Research (Aliso Viejo, Californien U.S.A)	Schlaufen zur Proteinkristallfixierung
Invitrogen (Karlsruhe)	<i>E. coli</i> -Stämme (TOP10, BL21 und BL21 Star) pBAD102/D-TOPO Expression Kit, pTrcHis2 TOPO® TA Expression Kit, Champion™ pET Directional TOPO® Expression Kit
<i>Iris Biotech</i> (Marktredwitz)	geschützte Aminosäuren
<i>Merck4Biosciences</i> (Nottingham, Großbritannien)	2-Chlorotriylchlorid-Harz, geschützte Aminosäuren, HBTU, HOBt, PyBOP
<i>Millipore</i> (Schwalbach)	Proteinkonzentratoren (Amicon Ultra-15)
<i>MP Biomedicals</i> (Illkirch, Frankreich)	Coenzym A Trilithiumsalz
<i>New England Biolabs</i> (a. M.)	Desoxyribonucleotide (dATP, dTTP, dGTP, dCTP), Restriktionsendonucleasen, 1kb-DNA-Ladder
<i>Nextal Biotechnologies</i> (Montreal, Kanada)	Crystall Screens: (Nextal Classic Suite, Nextal Classic Lite Suite, Nextal Anion Suite, Nextal MBC I&II Suite)
Operon, <i>Eurofins MWG</i> (Ebersberg), Sigma-Aldrich (München)	Oligonucleotide
<i>Oxoid</i> (Cambridge, Großbritannien)	Agar Nr. 1, Trypton
<i>Qiagen</i> (Hilden)	QIAprep spin Minprep Kit, QIAquick Gel Extraction Kit, QIAquick PCR Purification Kit, QIAGEN Genomic-tip 20/G, Ni <sup>2+</sup> -NTA-Agarose
Roth (Karlsruhe)	Ethydiumbromid, β-Mercaptoethanol, Acrylamid-Lsg. für SDS-PAGE
<i>Sarstedt</i> (Nümbrecht)	Pipettenspitzen, Spritzenversatzfilter Filtropur S0.2
<i>Nanoprobes</i> (Yaphank, USA)	Nanoprobes Labeling Reagenz
<i>Schleicher &amp; Schüll</i> (Dassel)	Sterilfilter (Porengr. 0.2 µm), Whatmann-3 <sup>MM</sup>

<i>Sigma-Aldrich</i> (Deisenhofen)	EDTA, -CoA Crystall Screens: (Cryst. Basic Kit, Cryst. Extension Kit, Cryst. Cryo Kit, Cryst. Basic Membrane Kit)
------------------------------------	--

Tabelle 8: Chemikalien, Enzyme und Verbrauchsmaterial

## 3.3 Verwendete Software und Datenbanken

Acronym/Programm	URL	Anwendungsgebiet
@AbsEdges	skuld.bmsc.washington.edu/scatter/AS_periodic.html	Absorptionsgrenzen Datenbank
@T-Coffee	www.tcoffee.org	Alignent von Protein- und DNA-Sequenzen
@SWISS-TOOLS	http://www.expasy.org	Berechnung verschiedener struktureller, physikalischer und chemischer Protein-Parameter
@Gimp	www.gimp.org/downloads	Bildbearbeitung
@Adobe Photoshop CS4	www.adobe.com/de	
@CloneManager	www.scied.com/pr_cmpro.htm	<i>In silico</i> -Klonierung
@Mozilla	http://www.mozilla-europe.org/de/	Internet Browser
@CCP4i	www.ccp4.ac.uk/ccp4i_main.php	Kristallographie Softwarepaket
@Wikipedia	http://de.wikipedia.org	Lexikon
@PDBsum	www.ebi.ac.uk/pdbsum	PDB Daten Aufarbeitung
@PM-Calculator	rna.rega.kuleuven.ac.be/masspec/pepcalc.htm	Petidmassen Rechner
@Peptide-Mass	expasy.org/tools/peptide-mass.html	
@Adobe Reader 9.0	www.adobe.com/de	Portable Document Format (PDF)
@PDF-XChange	pdf-xchange-viewer.softonic.de	
@EditSeq	www.dnastar.com/t-sub-products-lasergene-editseq.aspx	Primerdesign
@OligoCalc	www.basic.northwestern.edu/biotools/oligocalc.html	
@CLC Free Workbench	http://www.clcbio.com	Protein- und DNA-Alignmentsoftware
@Aline	crystal.bcs.uwa.edu.au/px/charlie/software/aline	
@Protein-Digest	db.systemsbio.net/proteomicsToolKit	Protein-Fragment Rechner
@SAXS	utilitieswww.szutucki.de/SAXSUtilities	SAXS Daten Prozessierung
@ATSAS	www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html	
@XtalPred	ffas.burnham.org/XtalPred-cgi/xtal.pl	Analyse und Vorhersage von Proteinstrukturen
@DSSP	srs.ebi.ac.uk/srsbin/cgi-bin/wgetz	Sekundärstruktur Datenbank
@PyMOL 1.3	www.pymol.org	
@Qutemol	qutemol.sourceforge.net	Visualisierungssoftware für Strukturdaten
@Swiss-pdb Viewer	spdbv.vital-it.ch	
@UCSF Chimera	www.cgl.ucsf.edu/chimera	
@X-ray Detector Software	xds.mpimf-heidelberg.mpg.de	Strukturdatenindexierung
@XSCALE	xds.mpimf-heidelberg.mpg.de/html_doc/xscale_program.html	Strukturdatenintergration
@DNA	www.dna.ac.uk	Synchrotron Datenerhebung
@Calc 3.2	de.openoffice.org	Tabellenkalkulation
@Excel 2007	office.microsoft.com/de-de/excel	
@Writer 2007	de.openoffice.org	Textbearbeitung
@Engauge	http://digitizer.sourceforge.net/	Translation von Bilddaten in tabellarische Dokumente
@Adobe Illustrator 10, CS3-4	www.adobe.com/de	
@Inkscape	www.inkscape.org	Vektorbasiertes Zeichenprogramm
@Coot	www.yzbl.york.ac.uk/~lohkamp/coot/	Verfeinerung von Proteinstrukturmodellen
@Leo	http://dict.leo.org	Wörterbuch englisch ↔ deutsch



Tabelle 9: Verwendete Software und Datenbanken

### 3.4 Verwendete Oligonukleotid-Primer

Primer	Funktion/Schnittstelle	Sequenz 5'→'	T <sub>m</sub> [°C]
AT#01	Punktmutations-Vorwärts- Primer für SrfAC S1003A	5'- gcgctcggaggcatgccttgaaggccatgaccg -3' Punktmutation Serin zu Alanin	78°C
AT#02	Punktmutations-Rückwärts- Primer für SrfAC S1003A	5'- cggatcatggccttcaaggatgcctccgagcgc -3' Punktmutation Serin zu Alanin	78°C
AT#07	Sequenzierungs- Primer SrfAC	5'-tcagctggcggaagaatg-3'	59°C
AT#37	Forward Primer: tycB23flf (BamHI Schnittstelle)	5'-aaaaaaca- ggatccggggaggaaagcgcgtatctcgccatt-3'	65°C
AT#28	Reverse Primer: tycB23flr (Xho I Schnittstelle)	5'-aaaaaa- ctcgagtcaaatccattccaggatgtttccagctcctc-3'	65°C
AT#31	Forward Primer: srfAA13f (BamHI Schnittstelle)	5'-aaaaaacaggatccggaaataac- ttttaccctttaacggatgcacaaaaacgaatttg-3'	65°C
AT#32	Reverse Primer: srfAA13r (Xho I Schnittstelle)	5'-aaaaaaactcgag- ttagaaaatttcattaatttatccagctcatccagcgtc -3'	65°C
AT#33	Forward Primer: srfAB13f (BamHI Schnittstelle)	5'-aaaaaacaggatcc-gagcaaaaaa- tcgattcaaaaggtgtacgcactgacac-3'	65°C
AT#34	Reverse Primer: srfAB13r (Xho I Schnittstelle)	5'-aaaaaaactcgagtatttt- aaattctcctaagcatgtcaagatatctccatct-3'	65°C
AT#43	Forward Primer: tycB23flf (pET Champion TOPO)	5'-caccggcggggaggaaa- gcgcgtatctcgccattcccc-3'	72°C
AT#44	Reverse Primer: tycB23flr (pET Champion TOPO)	5'-tcaaatccattccaggatgtttt- ccagctcctccatggacagctcttcac-3'	72°C
AT#58	Forward Primer: srfAAf (pET Champion TOPO)	5'-caccatgagaggatcgcatcaccatcaccat- cacggatccatggaaataacttttaccctttaacgg -3'	74°C
AT#59	Reverse Primer: srfAAr (pET Champion TOPO)	5'- gaaaatttcattaatttatccagctcatccagcgt- caaatcgtcgtcaccaagatcactcggggtc-3'	74°C

Tabelle 10: Verwendete Oligonukleotid-Primer (Schnittstellen sind unterstrichen).

Die Schmelztemperaturen (T<sub>m</sub>) für den wurden nach der „Nearest Neighbor“ Methode berechnet[141].

### 3.5 Vektor-Systeme

Die verwendeten Konstrukte wurden in die entsprechenden Vektoren kloniert (*srfAC* in pTrcHis2-TOPO- und *tycB*<sub>2-3FL</sub> in pET101/200D-TOPO- sowie im pCR-TOPO-XL-Vektor von *Invitrogen*) oder lagen bereits in dem pQE60-Vektor (*Qiagen*) kloniert vor (Abb. 18). Nach der Transformation in die entsprechenden Zellen erfolgte die Selektionskontrolle. Die Genexpression wird durch die Zugabe von Isopropyl-β-D-Thiogalaktosid (IPTG) induziert. Ausführliche Daten über den Promoter, MCS (multiple cloning site), die Resistenz, den Replikationsursprung sowie die Sequenz können den Datenblättern des Herstellers entnommen werden. Der pTrcHis2-TOPO-Vektor enthält am C-Terminus das *myc*-Epitop mit der Sequenz: -Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu-. Diese Sequenz wird vom Anti-*myc*-Antikörper (IgG<sub>1</sub>-Antikörper) erkannt, mit der konjugierten Meerrettich-Peroxidase und chemolumineszenten Substrat kann die Detektion und Analyse des Zielproteins in verschiedenen *Imaging*-Verfahren erfolgen[142].

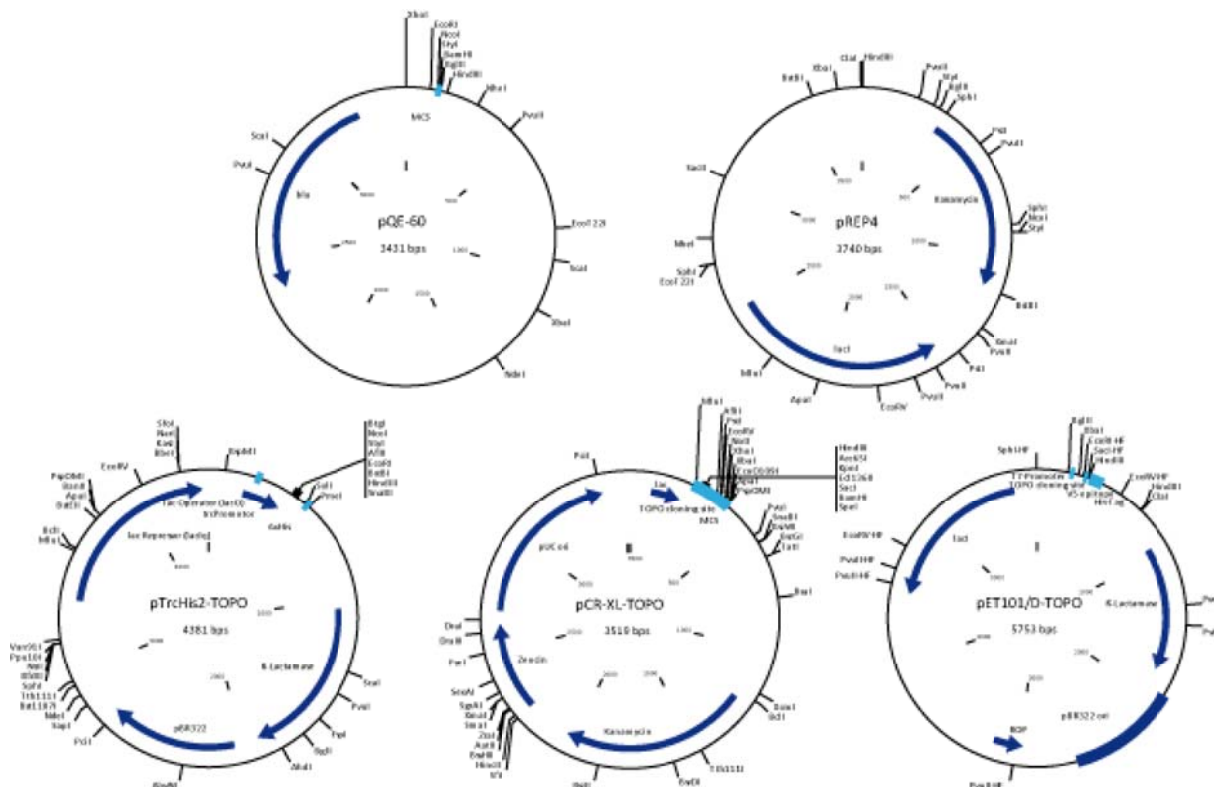


Abbildung 18: Vektorkarten für die verwendeten Plasmide: pQE60 mit dem Helferplasmid pREP4, sowie die Topoisomerase unterstützten Plasmide pTrcHis2, pCR-XL und pET101/200D.

### 3.6 Mikroorganismen

Stamm	Genotyp
<i>Escherichia Coli</i> Top10	<i>F-mcrA Δ(mrr- hsd RMS- mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 deoR araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG</i>
<i>Escherichia Coli</i> XL-1 Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Tetr)]</i>
<i>Escherichia Coli</i> BL21 (DE3)	<i>B F- ompT, hsdS (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>), dcm+ Tet<sup>r</sup>galλ(DE3) endA Hte</i>
<i>Escherichia Coli</i> BL21 Star	<i>B F- ompT hsdSB(r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) gal dcm rne131 (DE3)</i>
<i>Escherichia Coli</i> M15	<i>nalS, strS, rifS, thi-, lac-, ara+, gal+, mtl-, F-, rec+, uvr+, lon+</i>
<i>B. brevis</i> ATCC 8185	

Tabelle 11: Für Klonierung und Genexpression verwendete Mikroorganismen samt Genotyp.

### 3.7 Nährmedien

#### 3.7.1 LB-Medium

Die Anzucht von *Escherichia coli*-Stämmen erfolgt in zuvor autoklaviertem LB-Medium (*lysogeny broth*). Das LB-Medium wird aus in dH<sub>2</sub>O aufgelösten Bactotrypton (10 g/L ), Hefeextrakt (5 g/L) und NaCl (5 g/L) hergestellt. Anschließend wird der pH-Wert auf 7.0 eingestellt und das Nährmedium autoklaviert. Für Gel-Nährböden wird dem Medium davor Agar Nr. 1 in einer Endkonzentration von 1.2% (w/v) zugegeben. Nach Abkühlung des Festmediums auf mindestens 55°C kann das Festmedium mit Antibiotika versetzt werden.

#### 3.7.2 M9-Medium

Das M9-Medium ist ein Minimal-Wachstumsmedium, das für die Produktion von Selenomethionin-markierten Proteinen in *Escherichia coli*-Stämmen benutzt wird. Es besteht aus der Stammlösung und thermolabilen Zusätzen die sterilfiltriert dazugegeben werden.

Stammlösung 10xPUFFER:	K <sub>2</sub> HPO <sub>4</sub>	105 g/L
(autoklaviert)	KH <sub>2</sub> PO <sub>4</sub>	45 g/L
	tri-Natrium-Citrat	5 g/L

	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	10 g/L
CaCl <sub>2</sub> ( <i>autoklaviert</i> ):		100 g/L
MgSO <sub>4</sub> ( <i>autoklaviert</i> ):		100 g/L
Glucose ( <i>steril filtriert</i> ):		400 g/L
Biotin/Thiamin ( <i>steril filtriert</i> ) je:		2 g/L

Für 500 mL Expressionsvolumen an M9-Medium oder 50 mL Anzuchts-M9-Medium wurden die einzelnen Stammlösungen in folgendem Verhältnis vermengt:

	500 mL	50 mL
Stammlösung 10xPUFFER:	50 mL	5 mL
CaCl <sub>2</sub>	200 µL	20 µL
MgSO <sub>4</sub>	500 µL	50 µL
Glucose	5 mL	500 µL
Biotin/Thiamin	500 µL	50 µL
ddH <sub>2</sub> O ( <i>autoklaviert</i> ):	444 mL	44,4 mL

Für die Hemmung der zelleigenen Produktion an Methionin und deren Ersatz durch Selenomethionin wurde noch folgenden Stammlösungen gebraucht:

Aminosäure-Mix 1: L-Lys/L-Phe/L-Thr ( <i>steril filtriert</i> ) je:	10 g/L
Aminosäure-Mix 2: L-Ile/L-Leu/L-Val ( <i>steril filtriert</i> ) je:	5 g/L
Selenomethionin-Lösung ( <i>steril filtriert</i> ):	5 g/L

## 3.7.3 TB-Medium

Für die autoinduktive Expression von Vektoren, die einen Insert größer als 5 kb tragen, wurde das TB-Medium gewählt. Das TB-Medium enthält pro Liter:

Trypton	12 g
Hefeextrakt	24 g
Glycerin	4 ml oder 5 g
Kaliumdihydrogenphosphat	2.13 g
Dikaliumhydrogenphosphat	12.54 g

In das sterile Medium wurden anschließend 10 mL einer 40%-igen, sterilfiltrierten Laktose-lösung hinzugegeben.

## 4 Methoden

### 4.1 Molekularbiologische Methoden

Die Aufgabenstellung dieser Arbeit beschreibt die Projektkonzeption, anhand dessen die entsprechenden Zielproteine, sowie Zielgene ausgewählt und anschließend biochemisch sowie strukturell untersucht wurden. Die Klonierung der Zielgene erfolgt mittels Templat-angepasster Methode der Polymerasekettenreaktion (*polymerase chain reaction*, PCR) mit der Phusion Polymerase (*Finnzymes*)[143] oder für Genamplifikation von über 10 kb mit der DyNAzyme Polymerase (*Finnzymes*)[144]. Als Templat-DNA dient entweder die extrahierte Plasmid-DNA[145] bereits klonierter Konstrukte oder für das Zielgen als Wildtyp (WT) zuvor präparierte chromosomale DNA[146]. Um für die Untersuchungen relevante Variationen des Zielprotein-Wildtypes zu erhalten, werden die Zielgene nach der Klonierung, bzw. Ligation[147] in die entsprechenden Vektoren einer Dpn I-unterstützten Mutagenese-PCR[148] unterzogen. Die Ausgangsplasmide müssen dafür in der methylierten Form vorliegen und dürfen nicht aus einem Methyltransferase-defizienten Stamm (z.B. *E. coli* BL21) isoliert werden. Die Klonierung und Mutagenese werden nach dem *Vectormapping*-Verfahren analysiert, dafür werden Plasmide durch Restriktionsnukleasen (*NEB*)[149] verdaut und die Fragmente anschließend mittels Gelelektrophorese[150] visualisiert. Die Plasmid-DNA wurde anschließend sequenziert (*GATC*)[151]. Um die hergestellten Genkonstrukte optimal einzusetzen, wurde deren Plasmid-DNA extrahiert um durch eine Elektroporation in kompetente Zellen wieder transformiert werden zu können[152].

### 4.2 Proteinchemische Methoden

Zur Herstellung heterologer Proteine in *E. coli* werden die in den verwendeten Vektoren klonierte Gene in den *E. coli* - Stämmen M15, BL21 und BL21 Star[153] exprimiert. Durch Isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG) oder Laktose Zugabe erfolgt die Abspaltung des Repressors von der Plasmid-DNA und ermöglicht so die Genexpression des gewünschten Gens, zusätzlich kann die Bindung des Repressors vor der Induktion durch die Zugabe von

Glukose positiv beeinflusst werden.

Die Vektor-Systeme mit dem Zielgen werden zuerst in entsprechende kompetente Zellen transformiert und anschließend auf Agarplatten kultiviert. Für die Validierung der Ausbeuten werden die Einzelkolonien in analytischen Maßstab (50 mL) einer Testexpression unterzogen. Für die Vorkulturen wird das autoklavierte Nährmedium mit Kanamycin (25 – 50 µg/mL) und bzw. oder Ampicillin (50 – 100 µg/mL) versehen, mit Einzelkolonien inokuliert und über Nacht bei 30 – 37 °C und 250 upm inkubiert.

### 4.2.1 Genexpression im präparativen Maßstab

Die Konstrukte *tycB<sub>2-3</sub>* [AT.CATE], *tycB<sub>3</sub>* [CATE] und *tycB<sub>2-3</sub>ΔE* [AT.CAT][154] sowie *bacA<sub>1-2</sub>* [AT.CyAT][128] als auch die unimodularen Konstrukte *tycA* [ATE][155], *grsA* [ATE][155] bzw. *srfAC*[156] lagen im pQE60- bzw. im pTrcHis2-Vektor kloniert vor. Die klonierten Konstrukte *tycB<sub>2-3FL</sub>* [CAT.CATE] und *srfAA* [CAT.CAT.CATE] sowie das *lgrD<sub>1-2</sub>* [FAT.CAT][79] wurden im Autoinduktionsmedium (TB-Medium) exprimiert.

Die präparativen Expressionen (5 – 8 L Expressionsmedium) werden im Verhältnis 1:100 mit der Vorkultur inokuliert und bei 30 °C und 160 upm im Luftschüttler bis zu einer OD<sub>600</sub> von 0.6 inkubiert. Die Induktion der Expression erfolgt durch Zugabe einer IPTG-Lösung auf eine Endkonzentration von 1 mM IPTG. Die Kultur wird für weitere zwei bis vier Stunden bei 30 °C und 160 upm inkubiert. Danach werden die Zellen durch Zentrifugation geerntet (30 min, 6000 upm, 4° C) und das Zellpellet in 3%-Hepes-Puffer B (Hepes-Puffer A: 50 mM Hepes und 300 mM NaCl; Hepes-Puffer B: 50 mM Hepes, 300 mM NaCl und 250 mM Imidazol) resuspendiert. Bei Autoinduktionsmedium erfolgt die Induktion durch die enthaltene Laktose, IPTG wurde nicht zugegeben, die Zellkultur wurde nach zwölf bis achtzehn Stunden geerntet.

Die Vorkulturen sowie das Nährmedium für die präparative Expression Selenomethionin-markierter Proteine werden zusätzlich mit einer Glukose-Lösung (Endkonzentration von 22.2 mM) versetzt. Für die präparativen Expressionen werden 8 L Expressionsmedium im Verhältnis 1:400 mit der Vorkultur (M9-Medium) inokuliert und bei 30°C und 160 upm im Luftschüttler bis zu einer OD<sub>600</sub> von 0.5 inkubiert. Um den Einbau des zelleigenen Methionins zu minimieren, werden die Kultur anschließend mit dem Aminosäure-Mix 1 und 2 im

Verhältnis 1:100 versetzt und für ca. 15 min inkubiert. Im Anschluss daran wird die Kultur, um den Einbau des zelleigenen Methionins durch das Selenomethionin zu ersetzen, mit der Selenomethionin-Lösung, ebenfalls im Verhältnis 1:100, versetzt. Nach weiteren 15 min Inkubationszeit erfolgt durch Zugabe einer IPTG-Lösung auf eine Endkonzentration von 0.15 mM IPTG die Induktion der Zellkultur. Diese wird für weitere vier Stunden bei 30°C und 160 upm inkubiert, danach die Zellen durch Zentrifugation geerntet (15 min, 6000 upm, 4 °C) und das Zellpellet in 3%-Hepes-Puffer B resuspendiert. Alle Zellsuspensionen können direkt aufgeschlossen oder bis zur weiteren Verwendung bei -20°C gelagert werden.

#### 4.2.2 Zellaufschluß

Der Zellaufschluss wird mit einer French-Press (*Sim Aminco*) durchgeführt. Testexpressionen (50 µL Zellsuspension aus 1 – 2 mL Kulturmedium) wurden durch die Zugabe von Natriumdodecylsulfat (SDS) aufgeschlossen.

Bei dem Verfahren für die Expressionsansätze werden die Zellen zunächst einem hohen Druck ausgesetzt, der durch das Öffnen eines Nadelventils schlagartig abnimmt und zu einer rapiden Expansion der Zellen führt. Die dadurch auf die Zellwand wirkenden hohen Scherkräfte führen dabei zur Lyse der Zelle und zur Freisetzung des Zellinhaltes. Um einen vollständigen Zellaufschluss zu gewährleisten, wird der Vorgang mindestens dreimal unter kontinuierlicher Kühlung der Suspension wiederholt. Das erhaltene Lysat wird durch zweifache Zentrifugation (2x30 min, 17000 upm, 4°C) von unlöslichen Bestandteilen befreit und durch weiterführende Methoden aufgereinigt.

#### 4.2.3 Proteinaufreinigung

Die Metallchelate-Affinitätschromatographie an immobilisierten Metallionen (*immobilised metal affinity chromatography*; IMAC) macht sich die hohe Affinität von His<sub>x</sub>-Appendix-Proteinen an bivalenten Metallionen zunutze[157]. Hierbei nutzt man die Fähigkeit, Matrixgebundene Metallkationen wie die Cu<sup>2+</sup>-, Ni<sup>2+</sup>-, Zn<sup>2+</sup>- und Co<sup>2+</sup>-Ionen durch zwei Histidinreste zu komplexieren, dabei nimmt die Bindungsaffinität von Cu<sup>2+</sup> zu Co<sup>2+</sup> ab, die Spezifität nimmt gleichzeitig zu[158-159]. Die Trennung der rekombinanten Proteine von den übrigen löslichen Substanzen im Lysat der Expressionszellen erfolgt durch diese reversible Bindung an



der  $\text{Me}^{2+}$ -NTA-Matrix. Die Proteine werden über einen stabilen oktaedrischen Metalkation-Komplex an die Matrix gebunden und können von weiteren Proteinen des Zellrohextraktes isoliert werden. Die Elution der komplexierten Proteine erfolgt über eine ansteigende Imidazol-Konzentration im Elutionspuffer (Hepes B). Durch die höhere Affinität des Imidazols zu den Matrix-gebundenen Metallkationen werden die komplexierten Proteine von den Koordinationsstellen der Matrix verdrängt.

Zur Reinigung des Rohzellextraktes wurde eine FPLC-Anlage (*Pharmacia LKB Biotechnology*) oder eine Äkta Prime (*GE-Healthcare*) verwendet. Alle Reinigungsschritte werden mit gekühlten und filtrierten Puffern und bei einer Umgebungstemperatur von 4 °C durchgeführt. Zunächst wird der Rohzellextrakt auf eine Auftragsschleife gegeben und anschließend auf die  $\text{Me}^{2+}$ -NTA-Säule (*Pharmacia*) aufgetragen. Als Auftragspuffer diene ein 3% HEPES B-Puffer. Die von der Säule eluierten Proteine wurden durch einen UV-Licht-Detektor ( $\lambda=210\text{ nm}$ ) verfolgt, wobei das Messsignal und das Gradientprogramm auf einem xy-Schreiber ausgegeben werden. Durch einen Waschschrift mit 5% Hepes B-Puffer werden zuerst unspezifisch gebunden Proteine eluiert. Durch kontinuierliche Erhöhung der Hepes B-Anteils in dem Elutionsgradienten werden die gebundenen Proteine bei einer konstanten Flussrate durch das Imidazol von der Matrix verdrängt. Das Eluat wird anschließend durch einen automatisierten Fraktionskollektor in 2-4 mL-Fraktionen aufgefangen. Im Anschluß dient der qualitative Bradford-Assay[160] zur Bestimmung der proteinhaltigen Fraktionen. Die Analyse der positiv getesteten Fraktionen erfolgt mittels SDS-PAGE[161]. Die das Zielprotein enthaltenden Fraktionen werden anschließend vereinigt, in Amicon Ultra-15 Spin-Säulen (*Millipore*) ankonzentriert und weiter Verarbeitet, bzw. Analysiert.

### 4.2.4 Gelfiltrations-Chromatographie

Die Trennleistung der Gelfiltrationschromatographie beruht auf der unterschiedlichen Diffusionsgeschwindigkeit verschiedener Proteine in einer Gelmatrix definierter Porengröße. Das Diffusionsverhalten der Proteine hängt von ihrer räumlichen Ausdehnung ab, die in erster Näherung mit ihrer Masse korreliert. Die Gelmatrix, poröse Kügelchen mit einem 34  $\mu\text{m}$  langem Durchmesser, besteht aus quervernetzter Agarose, kovalent gebunden an Dextran (Sephadex). Sie verzögert die Diffusion der Partikel mit abnehmender Größe und

ermöglicht so eine effektive Trennung nach der Proteinmasse. Das Elutionsvolumen der aufgetragenen Proteine hängt exponentiell mit ihrer Masse zusammen. Aus diesem Grund ist eine Abschätzung des molekularen Gewichts der aufgetragenen Proteine, nach einer Kalibrierung der Gelfiltrationssäule möglich. Die benutzte *HiLoad 26/60 Superdex 200 prep grade* (GE Healthcare) Gelfiltrationssäule ist für die präparative Trennung der Proteine, deren Molekulargewicht zwischen 30 und 600 kDa liegt, geeignet.

Vor der präparativen Trennung der rekombinanter Proteine wird die Gelfiltrationssäule erst mit Wasser gewaschen (2x320 mL ddH<sub>2</sub>O) und anschließend mit Dialysepuffer (2x320 mL 10 mM HEPES pH 7.0 und 150 mM NaCl) equilibriert. Alle Puffer werden vor Gebrauch mit einem Porenfilter von Kleinstpartikeln befreit und entgast. Das nach der IMAC-Aufreinigung ankonzentrierte Proteingemisch (2-4 mL) wird auf die Säule aufgetragen und anschließend bei konstanter Flussrate (2.0 mL/min) eluiert. Das Eluat wird durch einen automatisierten Fraktionskollektor in 2-4 mL-Fraktionen aufgefangen. Im Anschluss dient der qualitative Bradford-Assay zur Bestimmung der proteinhaltigen Fraktionen[160]. Die Zusammensetzung der so positiv getesteten Fraktionen wird mittels SDS-PAGE[161] visualisiert. Die das Zielprotein enthaltenden Fraktionen werden vereinigt und ankonzentriert. Anschließend wird das dialysierte Proteinkonzentrat für die weitere Verwendung vorbereitet. Die Gelfiltrationssäule wird nach Gebrauch mit 80 mL NaOH (0.5 mM), Wasser (2x320 mL ddH<sub>2</sub>O) und verdünnter Ethanol-Lösung (2x320 mL 20% EtOH) gespült.

## 4.2.5 HI-TRAP Umpufferungsverfahren

Um die durch Metallchelate-Affinitätschromatographie erhaltenen Zielproteinlösungen von Salzen und noch vorhandenem Imidazol zu befreien, wurde neben der Gelfiltrations-Chromatographie die Entsalzung mittels HiTrap<sup>TM</sup>-Säulen (Amersham Biosciences) durchgeführt. Die Entsalzung der Lösung basiert auf dem Prinzip der Gelfiltrationschromatographie. Die Entsalzung wurde nach Herstellerangaben durchgeführt.

## 4.2.6 Ankonzentrieren von Proteinlösungen

Da die Konzentrationen der aufgereinigten und entsalzten Proteine in der Regel gering sind, wurden die Lösungen durch Viva-Spin-Säulen ankonzentriert (Amicon Ultra-15 Spin-Säulen

von *Millipore*). Bei diesem Verfahren wird Proteinlösung durch Zentrifugation bei 4000 upm durch eine Membran gedrückt. Wichtigste Kenngröße der Viva-Spin-Säulen ist deren Ausschlussvolumen, welches angibt, ab welcher Größe ein Protein die Membranporen nicht mehr durchdringen kann. Somit wird die Pufferlösung während der Zentrifugation durch die Membran dringen, wohingegen die Proteine zurückgehalten werden und deren Konzentration steigt. Ein Ausfällen oder Eintrocknen der Proteine ist dabei zu verhindern.

### 4.2.7 SDS-Polyacrylamidgelelektrophorese (SDS-PAGE)

Die SDS-Polyacrylamidgelelektrophorese (SDS-PAGE) ermöglicht die Trennung von Proteinen nach ihrer Größe durch das Anlegen eines elektrischen Feldes. Durch die variable Vernetzung innerhalb der polymeren Matrix aus Polyacrylamid wird ein molekularer Siebeffekt erreicht. Beschränkt wird die Qualität der Auftrennung durch die natürliche Breite der Banden, d.h. die Größe des Probenvolumens. Deshalb wurde die Elektrophorese mit einer Kombination aus zwei Gelen betrieben: dem Sammelgel und dem Trenngel[162].

Die Gelelektrophorese wird in kleinen Vertikalkammern (Mini-Protean 3) der Firma *Bio-Rad* durchgeführt. Nach Einsetzen des Gels wird die Kammer mit SDS-Laufpuffer gefüllt, wobei auch das Gel überschichtet und die Probentaschen mit Puffer gespült werden. Die Elektrophorese erfolgt bei 10 mA (Sammelgel, pH 6.8) bzw. 20-30 mA (Trenngel, pH 8.0) pro Gel. Zur Größenskalierung werden 4 µL eines Proteinmarkers aufgetragen. Für die Visualisierung der analytischen und präparativen Expressionen werden die Zellpellets aus 2 mL Kulturlösung in 75 µL TE-Puffer und 25 µL 4x-SDS-Probenpuffer resuspendiert und 10 min bei 95°C erhitzt, was zur Lyse der Zellen führt. Nach der anschließenden Zentrifugation werden 10-20 µL des Überstands auf das Gel aufgetragen. Proteinbanden im Polyacrylamidgel werden durch eine Behandlung mit Coomassie-Brilliant Blue R250 angefärbt. Dieser Farbstoff bindet unspezifisch an kationische und hydrophobe Seitenketten von Proteinen[163] und ermöglicht die Visualisierung der einzelnen Proteinbanden. Das Polyacrylamidgel wird dafür mit der Färbelösung bedeckt und bis zum Siedepunkt erhitzt. Die heiße Färbelösung wird danach bei RT auf dem Tischschüttler geschüttelt. Ungebundener Farbstoff wird anschließend durch zweimaliges Aufkochen in Wasser aus dem Gel gelöst. Im Anschluss wird die Hintergrundfärbung des Geles durch das Schütteln in heißer

Entfärberlösung entfernt.

#### 4.2.8 Proteinkonzentrationsbestimmung

##### *i.) Qualitative Proteindetektion nach Bradford*

Wechselwirkt der in der Bradford-Lösung[160] vorhandene Farbstoff mit Aminosäureseitenketten eines Proteins, findet eine Verschiebung des Absorptionsmaximums von rot ( $\lambda = 465$ ) zu blau ( $\lambda = 595$ ) statt. Diese Verschiebung ermöglicht die Detektion der proteinhaltigen Fraktionen mit dem bloßem Auge. Dafür werden 100  $\mu\text{L}$  der Lösung einer Fraktion mit 10  $\mu\text{L}$  der Bradford-Lösung versetzt.

##### *ii.) Spektrometrische Proteinkonzentrationsbestimmung*

Das Prinzip der Konzentrationsbestimmung von Proteinlösungen mittels Nanodrop folgt dem gleichen photometrischem Grundsatz wie bei der spektrometrischen Verfahren, der Absorption bei einer Wellenlänge von  $\lambda = 280 \text{ nm}$ . Die photometrischen Messungen werden aufgrund des geringen für die Messung benötigten Volumens mit einem Nanodrop-Gerät bestimmt.

Dafür wird das Gerät zuerst mit 1.5  $\mu\text{L}$  ddH<sub>2</sub>O initialisiert und anschließend mit 1.5  $\mu\text{L}$  Referenzpuffer kalibriert. Für die Messung werden dann mindestens 1.5  $\mu\text{L}$  Proteinlösung aufgetragen und aus der gemessenen Absorption die Proteinkonzentration berechnet. Die für die Messung notwendigen Extinktionskoeffizienten wurden mit dem Online-Tool Protparam[164] ermittelt.

##### *iii.) Proteinkonzentrationsbestimmung mittels Refraktometer*

Mittels eines digitalen Tischrefraktometers wird die Proteinkonzentration von Proteinproben an der Beamline X33 (SAXS) am EMBL (Hamburg) verifiziert. Dafür wird zuerst die Refraktometeroptik mit Wasser und Isopropanol gereinigt. Anschliessend wird der Probenpuffer als Referenz gemessen, der so enthaltenen Wert wird von dem Messwert der Proteinlösung abgezogen. Die enthaltene Differenz wird dann mit der geräteabhängigen Konstante ( $x = 4864$ ) multipliziert, was zu der Proteinkonzentration in mg/mL führt.

## 4.3 Biochemische Methoden

### 4.3.1 Adenylierungsaktivitätsmessungen durch den ATP/PP<sub>i</sub>-Test

Das Zielprotein mit einer oder mehreren A-Domänen wird für eine Adenylierungsaktivitätsmessung (ATP/PP<sub>i</sub>-Austausch) im Probenpuffer (50 mM NaCl, 2 mM MgCl<sub>2</sub>, 25 mM HEPES pH 8.0) vorgelegt. Anschließend werden der „kalte“ Reaktionsmix, bestehend aus 1 mM ATP, 1 mM DTT, 5 mM NaPP<sub>i</sub>, 10 mM entsprechender Aminosäure und 2 μM Enzym, sowie alle Kontrollen (ohne Enzy[165]m, ATP oder „heißem“ <sup>32</sup>PP<sub>i</sub>, je 500 μL) vorbereitet. Für die „heiße“ PP<sub>i</sub>-Lösung werden zuerst 5 μL der <sup>32</sup>PP<sub>i</sub> Stocklösung zu 1 mL einer 5 mM NaPP<sub>i</sub>-Lösung gegeben, um anschließend 5 μL dieser Lösung mit 15 mL Scintillationsflüssigkeit zu vereinen. Nach dem Zählen von Ereignissen radioaktiven Zerfalls von <sup>32</sup>P sollten 100000 cpm (*counts per minute*) erreicht werden, eventuell muss die „heiße“ PP<sub>i</sub>-Lösung verdünnt bzw. mit <sup>32</sup>PP<sub>i</sub> angereichert werden. Um die Reaktion zu beginnen, werden 100 μL der hergestellten „heißen“ <sup>32</sup>PP<sub>i</sub>-Lösung zu 500 μL der Reaktionsmix-Lösung pipettiert. Der Reaktionsmix wird anschließend 5 – 30 min bei 25 – 37 °C inkubiert. Um die Reaktion zu quenchen, werden 750 μL Aktivkohle-Suspension (muss während der Entnahme gerührt werden), bestehend aus 8.92 g NaPP<sub>i</sub> × 10 H<sub>2</sub>O, 10 mL HClO<sub>4</sub> (70%) und 3.2 g Aktivkohle, dem Reaktionsmix hinzugegeben und anschließend vortexiert. Der Reaktionsmix wird für 3 min bei 13000 upm pelletiert und der Überstand abgesaugt. Das Pellet wird in 750 μL Waschlösung (wie Aktivkohlelösung nur ohne Aktivkohle) resuspendiert und erneut zentrifugiert, der Überstand wird abgesaugt. Dieser Schritt wird wiederholt. Anschließend wird das Pellet erneut in 750 μL Waschlösung resuspendiert und dann in 15 mL Scintillationsflüssigkeit in ein Scintillationsgefäß gegeben. Vor der Scintillationsmessung wird die Lösung vortexiert um eine möglichst homogene Verteilung der Aktivkohle in der Scintillationsflüssigkeit zu gewährleisten.

### 4.3.2 PCP-Beladung mittels Panthoteinierungsreaktion

Mittels der 4'-Phosphopantethein-Transferase Sfp aus *B. subtilis* werden posttranslational alle PCP-Domänen eines NRPS-Systems von der Apo- in die aktive Holo-Form überführt. Dafür wird das zu beladende Enzym im Beladungspuffer (25 mM HEPES pH 7.0, 1 mM MgCl<sub>2</sub>

und 250 mM NaCl) aufgenommen. Anschließend wird das Sfp (10  $\mu$ M) sowie die fünffache Menge an CoA oder einem CoA-Derivat hinzugegeben. Der Mix wird dann für möglichst vollständige Beladung über Nacht bei 4 °C langsam gerührt. Nach der Beladung wird der Reaktionsmix zentrifugiert um ausgefallenes Protein abzutrennen. Der Reaktionsmix wird dann auf das gewünschte Volumen ankonzentriert.

## 4.3.3 Allosterische Inhibitionsassays

Für die allosterischen Inhibitionsassays wird das Akzeptorenzym 10  $\mu$ M im Reaktionspuffer, (10 mM HEPES pH 8.0 und 100 mM NaCl) vorgelegt. Des weiteren werden ATP (5 mM Endkonzentration, gepuffert),  $MgCl_2$  (25 mM Endkonzentration) und die korrespondierenden Aminosäuren (250  $\mu$ M Endkonzentration) hinzugegeben. Anschließend wird der Inhibitor (1:100 – 1:1000) aus einer 100 mM Vorratslösung in DMSO hinzugegeben. Die Reaktionslösung wird dann für 20 – 30 min bei RT inkubiert und der Assay durch die Zugabe des Donorenzyms (10  $\mu$ M) gestartet. Nach Ablauf der gewünschten Reaktionszeit wird die Reaktion durch die Zugabe von 10% Trifluoressigsäure (TFA) gestoppt, die präzipitierten Proteine mittels Zentrifugation abgetrennt und das Reaktionsgemisch mittels LCMS untersucht.

## 4.3.4 Gekoppelte Amino-CoA und Aminosäure Beladung

Eine weitere Alternative zu dem mit radioaktivem  $^{32}P$  gelabelten  $PP_i$  für die Aktivität der einzelnen A-Domänen ist die gekoppelte Beladung durch Amino-CoA und der korrespondierenden Aminosäure. Ein Vorteil dieser Anwendung ist, dass sich die Aktivität mehrerer A-Domänen gleichzeitig untersuchen lässt, auch wenn mehrere die gleiche Aminosäure aktivieren. Zusätzlich Wie bereits beschrieben wird das Amino-CoA mittels Sfp auf die PCP Domäne beladen, zusätzlich wird ATP (100 eq) und für jede A-Domäne des Zielproteins die korrespondierende Aminosäure (250 eq) zugegeben. Die Lösung wird dann für präparative Zwecke bei 4 °C über Nacht langsam gerührt und für analytische Zwecke bei 20 °C für zwei Stunden langsam geschüttelt.

## 4.4 Bioanalytische Methoden

### 4.4.1 MALDI-TOF

Das Verfahren MALDI-TOF[166] dient zur analytischen Untersuchung von chemischen Verbindungen anhand ihrer Molekularmassen. Bei MALDI (*matrix assisted laser desorption/ionisation*) ionisiert man den Analyten mittels Laserbeschuss über eine DHB-Matrix (*Agilent*), in welche die eigentlichen Analyten eingebettet sind. Dafür werden 1 – 2 µL der Probe auf einem Metaltarget vorgelegt, mit 1 – 2 µL der DHB-Matrix vereint und anschliessend getrocknet. Durch Laserbeschuss verdampft die Matrix explosionsartig, das zuuntersuchende Analyt wird mitgerissen und ionisiert. Die entstandenen Ionen werden in einem elektrischen Feld beschleunigt. Die Masse des Analyts wird dann durch den TOF-Analysator (*time of flight*) über eine Zeitmessung vom Start der Ionen in der Ionenquelle bis zum Erreichen des Detektors über einen Deflektor, der für genaue Messergebnisse sorgt, ermittelt. Mit dieser Methode werden durch den BiFlex III MALDI-TOF-System (*Bruker*) hauptsächlich die Intermediate bei der Aufreinigung der Peptidsynthese-Produkte analysiert.

### 4.4.2 Hybrid-Massenspektrometrie (HPLC- und LC-MS)

Durch die Hochleistungs-Flüssigkeitschromatographie (*high performance liquid chromatography*; HPLC)[167] können anhand ihrer Affinität zum Säulenmaterial einzelne Substanzen oder Substanzgruppen aus Gemischen isoliert werden. An der unpolaren stationären Phase (*reversed phase*; RP) werden durch die polare mobile Phase (deionisiertes Wasser funktionalisiert mit 1% Trichloressigsäure als Ionenpaarreagenz) zuerst die polaren Stoffe eluiert. Anschließend werden durch einen Gradienten gegen eine unpolare mobile Phase (Methanol oder Acetonitril, ebenfalls mit 1% Trichloressigsäure) die adsorbierten Substanzen vom Säulenmaterial verdrängt und im Fraktionskollektor gesammelt.

Die Elution bei der präparativen HPLC-Methode wird über UV-Vis-Detektion verfolgt und die Fraktionen mit einem UV-Ausschlag (*peak*) mittels MALDI-TOF analysiert. Für die Identifikation von Proteinen werden die Peptidfragmente nach einem tryptischen Verdau mittels der gekoppelten HPLC- und MS-Analyse[168] mit einer ESI-Ionenquelle (*electrospray ionisation*)[169] an einem Agilent 1100 HPLC (*Agilent Technologies*) und Finnigan LTQ-FT



(*Thermo Fischer Scientific*) Hybridmassenspektrometer untersucht. Zur Verbesserung der Ionisierung der Analyten wird das Laufmittel durch die Zugabe von Nonafluorpentansäure (20 mM), Trifluoressigsäure (0.1%) oder Triethylamin (2 mM) als Ionenpaarreagenz funktionalisiert.

Aktivitätsassays werden mittels LCMS (*liquid chromatography mass spectrometry*)[170] analysiert, somit ist für jeden Punkt des gemessenen Chromatogramms ein Massenspektrum verfügbar. Dadurch können alle im Chromatogramm identifizierten Peaks einer spezifischen Substanz zugewiesen werden und qualitative sowie quantitative Aussagen über den analysierten Assay getroffen werden.

## 4.4.3 Bottom-Up Proteinanalyse

Eine der wichtigsten Disziplinen der Bioanalytik ist die Identifizierung von Proteinen sowie ihrer funktionalisierter Aminosäuren. Abhängig von der Anforderung können verschiedene Methoden zur Proteinanalyse (Abb. 19) gewählt werden. Geht man von einem intakten Protein aus, das in der Ionenfalle des Gerätes anschließend fragmentiert und mittels MS/MS untersucht wird, spricht man von der so genannten Top-Down Proteinanalyse. Eines der meistbenutzten Verfahren ist die Peptidmassen-Fingerabdruck Methode, bei der zuerst das Zielprotein mittels einer Protease, meist Trypsin, in Peptide gespalten wird, die anschließend durch eine MS/MS Einheit untersucht werden. Danach führt der

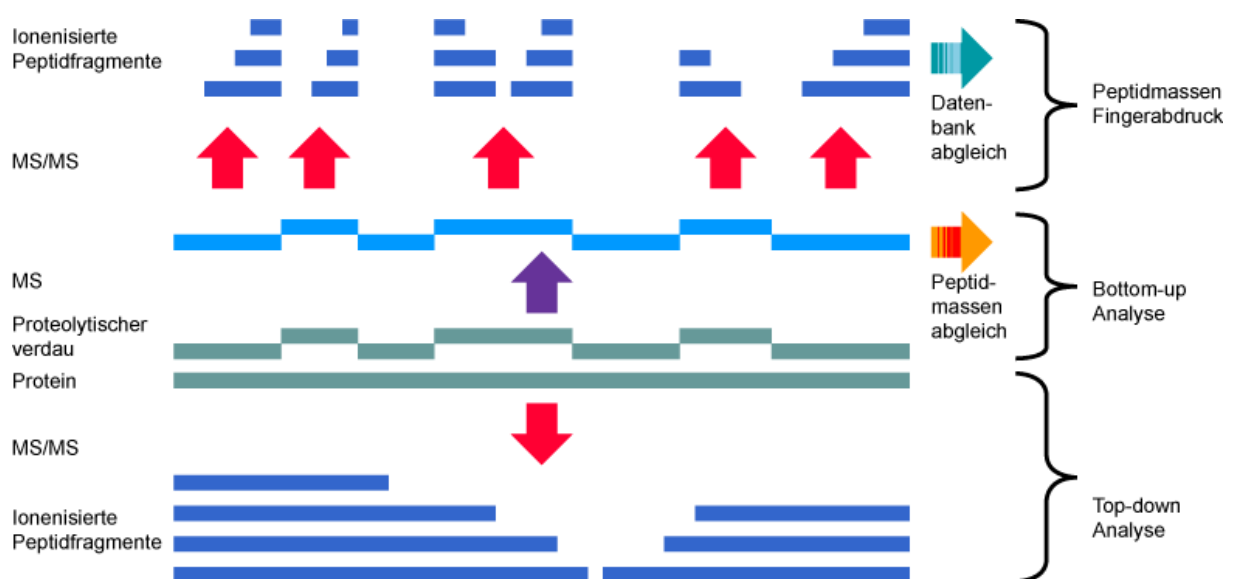


Abbildung 19: Massenspektrometrische Verfahren zur Proteinanalyse

Datenbankabgleich gegen alle Proteine des Organismus zum gesuchten Protein. In meisten Fällen ist die Proteinsequenz ganz oder zumindestens teilweise bekannt, was die anschließende Analyse stark vereinfacht. Für diese Zwecke ist die in dieser Arbeit verwendete Bottom-Up Proteinanalyse[171] geeignet, dabei werden die Peptidfragmente mittels MS analysiert und mit berechneten Werten verglichen.

### 4.5 Chemische Synthesen

#### 4.5.1 Peptidfestphasensynthese

Bei der Festphasensynthese (*solid phase peptide synthesis*, SPPS)[172] wird die Peptidsynthese an einem festen Träger, einem Polymer (*resin*), hergestellt. Die Immobilisierung am festen Träger bringt viele Vorteile mit sich. Zum einen können alle Reagenzien in grossen Überschuss eingesetzt sowie schnell von der festen Phase entfernt werden, was die Automatisierung der Festphasensynthese ermöglicht und die Kupplungsausbeuten deutlich steigert. Das Prinzip der Fmoc-basierten SPPS beruht auf der Orthogonalität von Schutzgruppen. So wird die basenlabile Fmoc-Schutzgruppe am N-Terminus der Aminosäuren und säurelabile Schutzgruppen (*acid labile protecting group*) für die Schützung der Seitenketten verwendet (Abb. 20).

Die Initiation der Peptidsynthese beginnt mit der Funktionalisierung des Harzes mit der ersten Aminosäure, bzw. an den auf der Harzoberfläche gebundenen 2-Chlorotriylchlorid-Linker. Dies kann im Vorfeld manuell erfolgen oder während der Festphasensynthese am Synthetiser. Für die Kupplung der nächsten Aminosäure muss der N-Terminus basisch entschützt werden. Nach mehreren Waschschritten erfolgt die Kupplung des nächsten Bausteines, der *in situ* mittels HBTU am C-Terminus für die Peptidbindungsknüpfung aktiviert wird. Nach mehrfacher Wiederholung (n-Schritte) dieser Propagationsschritte erfolgt die Kupplung der letzten Aminosäure. Die letzte Aminosäure verfügt nicht über die basenlabile Fmoc-Schutzgruppe, sondern tBoc-Schutzgruppe. Dies ist notwendig, da die Thioester/Peptid-Kupplung im nächsten Schritt unter basischen Bedingungen durchgeführt wird. Erst anschließend sollte der N-Terminus entschützt werden. Die Abspaltung des vollständig Seitenketten- und N-terminal-geschützten Peptids vom 2-Chlorotriylchlorid-Linker wird

durch Zugabe eines Gemisches aus AcOH/TFE/DCM erreicht. Unter diesen leicht sauren Bedingungen erfolgt die Abspaltung vom Harz. Die basenlabilen Seitenketten-Schutzgruppen (Fmoc) bleiben erhalten und werden erst im letzten Schritt entfernt um Nebenreaktionen mit den Aminosäure-Seitenketten zu vermeiden. Nach dem Abtrennen des Peptids vom Harz erfolgt die Reinigung durch dreifache Umkristallisation in Hexan und anschließend eine Aufreinigung mittels HPLC.

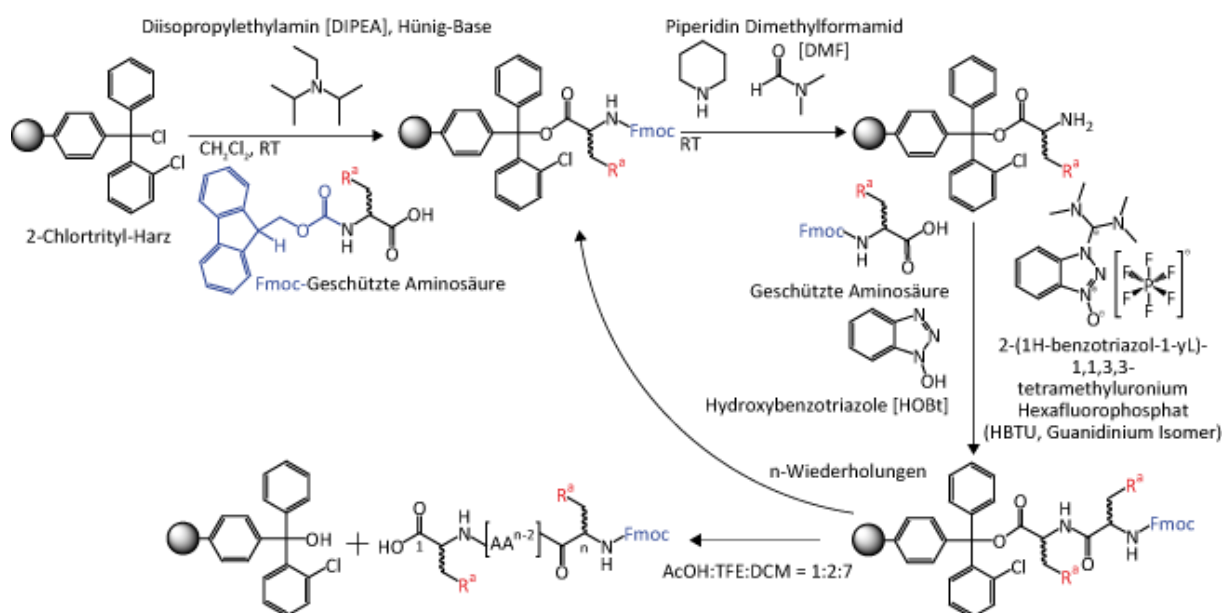


Abbildung 20: Peptid-Festphasensynthese (Fmoc-Strategie) mit der Fmoc-Schutzgruppe (blau) und der durch die säurelabile Schutzgruppe funktionalisierte Seitenkette der entsprechenden Aminosäuren ( $R^a$ , rot).

## 4.5.2 Amino-CoA Synthese

Da die Thioester Bindung zwischen der Thiolgruppe des Coenzym A (CoA) und der Carboxylgruppe des Substrates im Basischen zu Hydrolysierung neigt, wird diese durch eine Aminogruppe substituiert[173]. Zur Darstellung von Pantothenensäure (2) wird zuerst 1 eq des in Dichlormethan unlöslichem Pantothenensäure-Calciumsalz (1) in 200 mL  $\text{dH}_2\text{O}$  gelöst (Abb. 21). Um das Calciumsalz in die Säure zu überführen wird die Reaktionslösung mit 4M Salzsäure auf einen pH-Wert von 3 angesäuert, die Reaktionslösung im Vakuum eingeeengt und anschließend mit Ethylacetat mehrfach ausgeschüttelt. Das Ethylacetat der separierten organischen Phase wird im Vakuum entfernt. Für die Acetalschätzung der Pantothenensäure (2)

wird diese in 100 mL Dichlormethan (DCM) aufgenommen und anschließend das p-Anisaldehyddimethylacetal (PADA; 1 eq) zur Emulsion gegeben. Nach Zugabe von ( $\pm$ )-Campher-10-sulfonsäure (CSA; 0.1 eq) tritt eine zügige Vermischung der Substrate ein, was in einer klaren Reaktionslösung resultiert. Die Reaktion findet unter ständigem Rühren über Nacht statt. Nach Entfernen des Lösungsmittels im Vakuum wird das im Rückstand enthaltende Produkt (3) säulenchromatographisch über Silicagel (n-Hexan:Ethylacetat = 6:1; gefolgt von steigender Ethylacetat Konzentration bis hin zu reinem Ethylacetat; alle Lösungsmittel wurden mit 0.1% TFA angesäuert) von den restlichen Substanzen separiert. Der Fortschritt der säulenchromatographischen Separation wird mittels Dünnschichtchromatographie (DC) verfolgt. Nach Entfernen des Lösungsmittels im Vakuum enthält der weißliche Rückstand die acetalgeschützte Pantothensäure (3; Abb. 21).

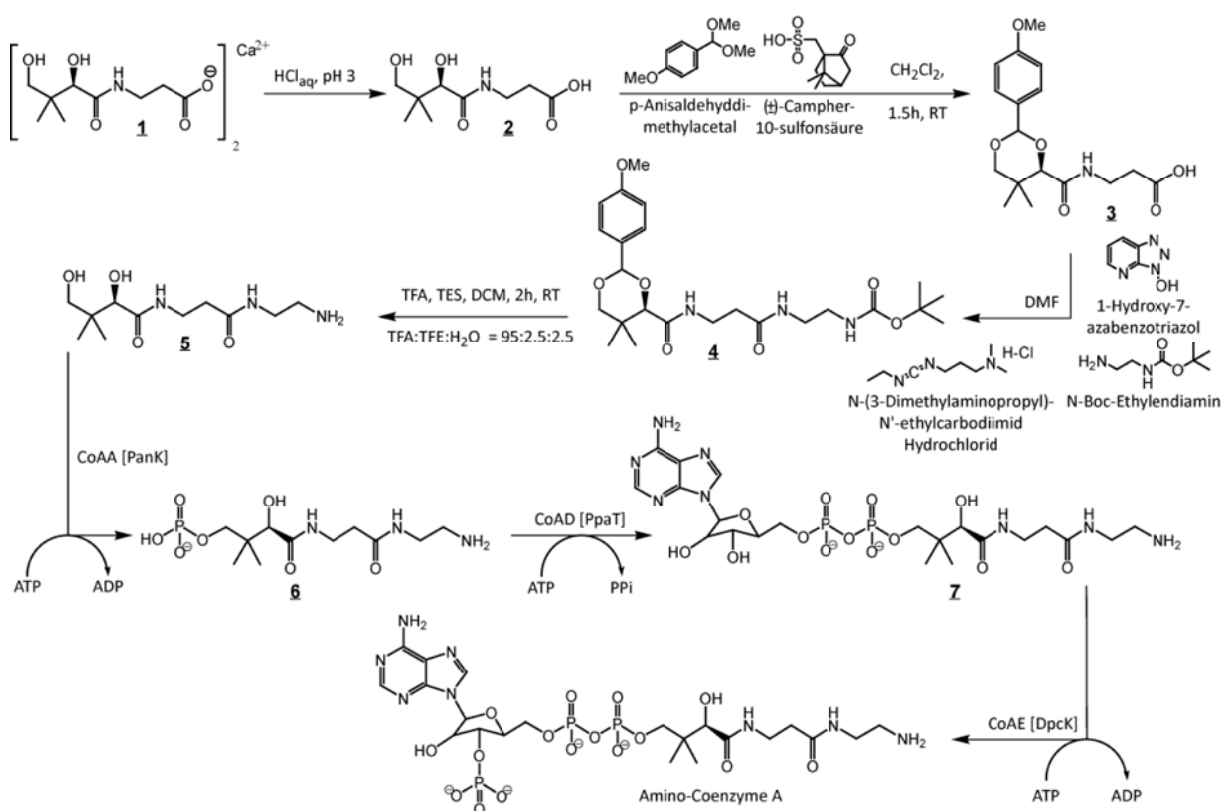


Abbildung 21: Chemoenzymatische Darstellung von Amino-CoA

Die Herstellung des Aminopantetheins aus der Pantothensäure erfolgt durch eine Peptidkupplung zwischen der Carboxygruppe der Pantothensäure und der freien Aminogruppe des N-Boc-Ethylendiamin (N-Boc-EDA; 1 eq). Dafür wird das Edukt (3) zusammen mit N-Boc-EDA, N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimid Hydrochlorid

(EDC; 2 eq) und 1-Hydroxy-7-azabenzotriazol (HOAt; 2 eq) in 15 mL Dimethylformamid (DMF) aufgenommen. Unter starkem Rühren wird das Triethylamin (NEt<sub>3</sub>; 3 eq) zu der Reaktionslösung langsam hinzugefügt, anschließend findet die Reaktion über Nacht unter ständigem Rühren statt. Nach dem Entfernen des Lösungsmittels im Vakuum (bei 40° C, 5 mbar für 4 h) kann das geschützte Aminopantethein (4) durch Ausschütteln aus 30 mL Ethylacetat und 20 mL gesättigter NaHCO<sub>3</sub>-Lösung aus der organischen Phase extrahiert werden. Die organische Phase wird mehrmals gewaschen (mit gesättigter NaHCO<sub>3</sub>-Lösung und zweimalig mit gesättigter NaCl-Lösung) und das Lösungsmittel im Vakuum entfernt. Durch säulenchromatographische Aufreinigung über Silicagel (6:1 n-Hexan:Ethylacetat; gefolgt von steigender Ethylacetat Konzentration bis hin zu reinem Ethylacetat; alle Lösungsmittel wurden mit 0.1% TFA angesäuert) erhält man das geschützte Aminopantethein (4; Abb. 21) als weißes Pulver. Anschließend werden die säurelabilen Schutzgruppen entfernt, zuerst für 4 h und Raumtemperatur (RT) mittels Triethylsilan (TES; 5 eq) und Trifluoressigsäure (TFA; 10 eq), die Kontrolle erfolgt mittels LCMS. Bei unvollständiger Abspaltung der Schutzgruppen wird nach der letzten Umsetzung die Entschützung durch ein Gemisch, bestehend aus Trifluoressigsäure (TFA) zu Trifluoroethanol (TFE) zu Wasser im Verhältniss von 95:2.5:2.5, durchgeführt. Um das finale Produkt Amino-Coenzym A (CoA-NH<sub>2</sub>) zu erhalten wird das Aminopantethein (5; Abb. 21) enzymatisch mittels Pantothenatkinase (PanK) an der terminalen Hydroxylgruppe phosphoryliert, das Pantethein-Phosphat-Adenylyltransferase (PpaT) adenosyliert die angehängte Phosphatgruppe und die Dephospho-CoA-Kinase (Dpck) phosphoryliert die 2-Hydroxylgruppe des Riboserestes. Die enzymatische Umsetzung zuerst durch PanK (0.0025 eq) und PpaT (0.0025 eq) erfolgt im Tris/HCl-Puffer bei einem pH-Wert von 7.0, Adenosintriphosphat (ATP; 10 eq), Magnesiumchlorid (MgCl<sub>2</sub>; 20 eq) sowie Kaliumchlorid (KCl; 20 eq). Anschließend, um das vollständig phosphorylierte Produkt zu erhalten, wird der pH-Wert auf 9.0 angehoben und Dpck (0.005 eq) hinzugeben. Das Reaktionsgemisch wird in einem 15 mL verschließbaren Reaktionsgefäß für 2 h bei RT geschüttelt. Da auch das dephosphorylierte Produkt (DP-CoA-NH<sub>2</sub>) von einer 4'-Phosphopantetheinyl Transferase wie Sfp übertragen wird, kann der letzte enzymatische Schritt der Einfachheit halber auch ausgelassen werden. Nach Ablauf der Reaktionszeit werden die Salze sowie die Enzyme mittels einer Sep-Pak® Plus C18

cartridge 55-105  $\mu\text{m}$  (Waters, WAT020515) entfernt. Dafür muss die Säule zuerst mit jeweils 10 mL Acetonitril und Wasser equilibriert werden, um im Anschluss das Reaktionsgemisch aufzutragen. Durch Waschen mit 5 mL Wasser werden die Salze ausgewaschen, die Elution erfolgt durch den Auftrag von 6 mL 0.1% TFA haltigen Gemisch (Acetonitril : ddH<sub>2</sub>O = 1 : 1). Das Produkt wird lyophilisiert und mittels HPLC aufgereinigt (Flussrate = 18 mL/min; C18-Säule; Laufmittel A: ddH<sub>2</sub>O mit 0,1% TFA; Laufmittel B: Acetonitril mit 0,1% TFA). Nach Lyophilisation der entsprechenden Fraktionen wird das Amino-CoA bei -80° C gelagert.

## 4.6 Methoden zur Aufklärung makromolekularer Strukturen

### 4.6.1 Dynamische Lichtstreuung (DLS) unterstützte Kleinwinkelstreuung (SAXS)

Die Kombination der Daten aus Experimenten der dynamischen Lichtstreuung (*dynamic light scattering*; DLS) und den aus der Kleinwinkel-Streuung (*small angle x-ray scattering*; SAXS) wird benutzt um die globale Struktur von heterogenen Makromolekül-Gemischen oder den zur Agglomeration neigenden Makromolekülen zu ermitteln[174].

Die dynamische Lichtstreuung ermöglicht anhand der Konzentrationsfluktuationen der Partikel im Probenpuffer (Brown'sche Molekularbewegungen) den hydrodynamischen Radius der Moleküle zu ermitteln[175]. Dabei wird mittels eines Lasers Licht in die Probe eingestrahlt das an verschiedenen Partikeln gestreut wird (Abb. 22).

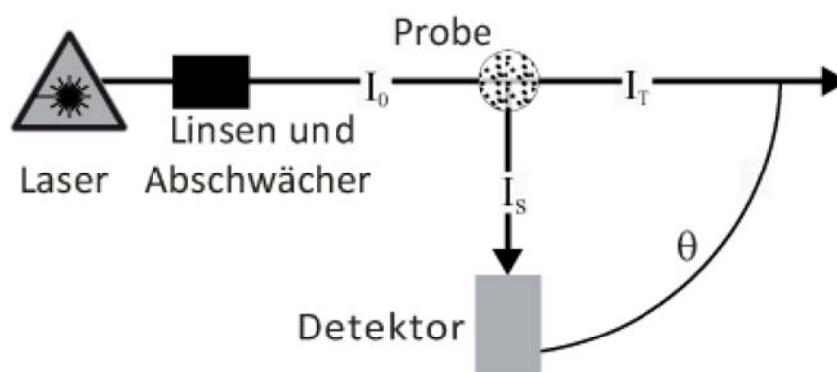


Abbildung 22: Schematischer Aufbau eines Lichtstreugerätes

Bei kohärent und monochromatisch eingestrahlttem Licht (Laser) interferiert das Streulicht verschiedener Streuzentren miteinander und führt zu messbaren Veränderungen in der Streuintensität. Die Fluktuationen in der Streuintensität enthalten Informationen über die

Geschwindigkeit, mit der sich die Partikel in der Lösung bewegen, die von dem Stokes'schem Radius der Partikel abhängt. Die Geschwindigkeit mit der sich die Partikel in der Lösung bewegen hängt von der stoffspezifischem Diffusionskoeffizient ab, mit dessen Hilfe sich mittels der Stokes-Einstein-Beziehung der hydrodynamische Radius  $R_0$  (Stokes'sche Radius) berechnen lässt.

Bei dem Streuexperiment (SAXS)[176-180] wird auf eine Lösung mit bekannter Proteinkonzentration Röntgenstrahlung bekannter Wellenlänge ( $\lambda$ ) eingestrahlt (Abb. 23). Die Streuintensität ( $I(q)$ ) wird als Funktion des Streuvektors  $q$  aufgezeichnet ( $q = 4\pi\sin\theta/\lambda$ ; wobei  $2\theta$  der Winkel zwischen der einfallenden und gestreuten Strahlung darstellt).

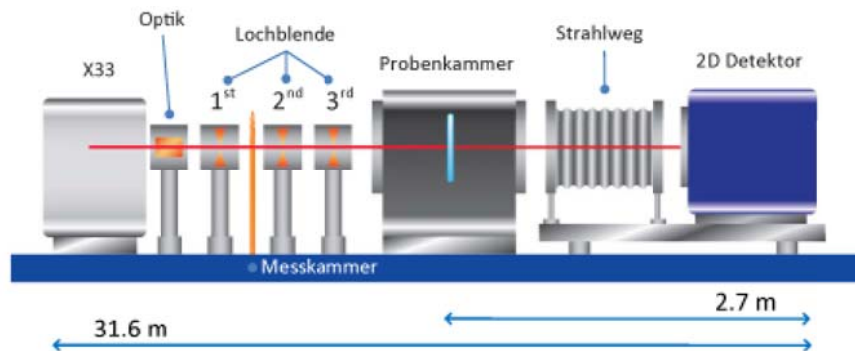


Abbildung 23: Schematischer Aufbau der Bio-SAXS-Messkammer (X33, HASYLAB)

Von so erhaltenen Daten werden dann die Werte des Lösungsmittels abgezogen (daher ist es von größter Wichtigkeit für die Integrität der Messung, dass sich bei dem Proben- und dem Kontrollpuffer um den identischen Puffer handelt). Die Auflösung beläuft sich bei einem Detektorabstand von 2.7 m auf  $q = 0.06 - 6 \text{ nm}^{-1}$ . Die Zielmoleküle in dem Probenpuffer haben eine zufällige Position sowie Orientierung, sofern sie monodispers vorliegen und nicht miteinander interagieren. In dem Fall kann eine isotropische Verteilung der Intensität gemessen werden, die proportional zu der Streuung eines einzelnen Moleküls, gemittelt über alle Richtungen ist. Bei einem Zwei-Phasen-System, so wie bei Proteinen in Lösung, wird der einfallende Hauptstrahl an dem Dichtekontrastunterschied ( $\Delta\rho$ ) zwischen der durchschnittlichen Elektronendichte des Partikels ( $\sim 0.44 \text{ e}^-/\text{\AA}^3$  für Proteine) und des Lösungsmittels ( $\sim 0.33 \text{ e}^-/\text{\AA}^3$  für Wasser) gestreut[180]. Zusätze im Puffer wie Chloridionen, ATP, DTT oder Glycerol erhöhen die durchschnittliche Elektronendichte in dem Puffer und verringern daher den Kontrast. Konventionell können die Strukturparameter: Originalintensität ( $I_0$ ) und



Gyrationsradius ( $R_G$ ) aus den gemessenen Intensitäten ermittelt werden. Die durch DLS-Messungen ermittelte Geschwindigkeit der Partikel führt zu dem Stokes'sche Radius ( $R_0$ ), der für den Radius einer idealisierten Sphere, die sich mit der gleichen Geschwindigkeit durch die Lösung bewegt wie das gemessene Partikel steht. Der Gyrationsradius  $R_G$  wird als die Quadratwurzel der mittleren Quadrate aller Distanzen der Streuzentren eines Partikels von seinem Zentrum definiert und hängt somit ab von der Form des gemessenen Partikels ab[181]. Wird der Logarithmus der Streuintensität ( $\ln I(q)$ ) gegen den Quadrat des Streuvektors ( $q^2$ ) aufgetragen erhält man den Guinier-Plot[181]. Aus dem Guinier-Plot lässt sich der Gyrationsradius ( $R_G$ ) als Steigung aus der Extrapolation für den Bereich  $q < 1.3$  (Guinier Näherung) gegen  $I(0)$  ( $q = 0$ ) ermitteln[180]. Durch das Porod-Gesetz sind die Grenzen der Kleinwinkel-Streuung vorgegeben, da für sehr große Streuvektoren ( $q \gg 1$ ) die Streuintensität mit der vierten Potenz des Streuvektors abfällt[182]. Die Streukurve hängt dann mit zunehmendem Streuvektor stärker von der Form der Teilchen ab, ist aber auch immer anfälliger für Messfehler, da die Streuintensität im Hintergrundrauschen verschwindet. Dies eröffnet aber auch die Möglichkeit Aussagen über die Form der gemessenen Teilchen zu treffen. Aus der fouriertransformierten Streuintensität führt die gesuchte Dichteverteilung  $\rho(r)$  (Paarverteilungsfunktion; engl. *pair distribution function*). Aus dieser lassen sich erste Strukturinformationen des gesuchten Proteins berechnen. Ausgehend von der Dichteverteilungsfunktion werden *ab initio* Strukturmodelle berechnet, aus denen wiederum Modell-Streukurven berechnet werden. Die *ab initio* Strukturmodelle werden so angepasst, dass die Abweichung zwischen der berechneten und der gemessenen Streukurve minimiert werden. Da die zweidimensionalen Streukurven als mittlere Verteilung aller Partikel und Konformationen gemessen werden, können mehrere *ab initio* Strukturmodelle die minimale Abweichung zu der gemessenen Streukurve zeigen. Die erstellten Modelle werden klassifiziert und anhand der strukturellen Merkmale verfeinert. Wie gut das Modell verfeinert wurde zeigt die Größe für die Anpassungsqualität  $\chi^2$  an. Ein Wert von 1 zeigt eine vollständige Übereinstimmung, die jedoch praktisch nicht zu erreichen ist, ein Wert im Bereich zwischen 1.3 und 2.0, je nach Qualität der Daten, ist daher eher realistisch.

*i.) Datensammlung*

Die SAXS Datensätze wurden an der X33 Beamline (EMBL) des DORIS III Rings (*storage ring*) am Deutschen Elektronen Synchrotron (DESY) in Hamburg aufgenommen[183]. Die X33 Beamline (*bending magnet beamline*) hat eine feste Wellenlänge ( $\lambda = 1.5 \text{ \AA}$ ) und ist optimiert für die Aufnahme von Datensätzen mit niedrigem Hintergrundrauschen. Die Proben wurden durch einen automatischen Probenwechsler in die Messvorrichtung geladen[184]. Die Streubilder wurden mittels des PILATES 1M (*photon counting*) Detektors mit 4 x 30 s Bildern (*frames*) und für die Konzentrationspanne für 1 bis 9 mg/mL Proteinlösung mittels des *MAR Image Plate* – Detektors ( $67 \times 420 \text{ mm}^2$ ) mit einer 2 x 3 min Beleuchtungszeit aufgenommen[185]. Die Streukurven wurden durch AutoSub auf die Datenqualität analysiert[186] und während der Messung durch den Beamline Meta-Server koordiniert und überwacht[187].

*ii.) Datenreduktion und Analyse*

Die AutoSub Software korrigiert die Rückmeldung des Detektors, mittelt und überprüft die gemessenen Daten auf Strahlungsschäden. Mittels AutoRG[187] wurde der Gyrationradius ( $R_G$ ) ermittelt und durch AutoGNOM[188] die maximale Partikelausdehnung (DMAX). Streukurven für alle Konzentrationsbereiche wurden mit AutoMerge (ATSAS 2.3) auf ein Minimum (*zero concetration*) extrapoliert um alle vorhandenen Konzentrationseffekte im Anfangsbereich der Streukurven zu eliminieren. Die Messwerte der verschiedenen Konzentrationen wurden durch PRIMUS[189] integriert. Die Streukurvenselektion erfolgte nach der Superposition durch die Methode der kleinsten Quadrate und basiert auf der Anpassungsqualität ( $\chi^2$ ) mit einem Schwellenwert von 1.5.

*iii.) Modelaufbau und Verifikation*

Niedrig-Auflösung-Formmodelle (*low resolution shape model*) wurden durch DAMMIF[190] berechnet. Es benutzt eine Anordnung der Muster-Atome (*dummy atoms*) in einem kubischem Gitter, die die geringste Abweichung zwischen den prozessierten Messwerten und den errechneten Streukurven für die Muster-Atome nach der Kugelflächenfunktion (*spherical harmonics expansion*) zeigen. Für jeden zusammengefügt Datensatz werden 20 verschiedene Simulationsdurchgänge berechnet und anschließend analysiert. Um

Datenartefakte und messbedingte Fehler ausschließen zu können, wird für jede konzentrationsabhängige Messung ein Modell erstellt und visuell auf Übereinstimmungen und Abweichungen abgestimmt. Das Modell wird anschließend durch CRY SOL[191] nochmals verfeinert und mit einer Hydratationshülle versehen. Nach der Erstellung des Modells wird der Modellsatz mit dem Programm SASREF[192] durch die Domänenmodelle (*full atomic rigid domains models*) ergänzt. Durch den Vergleich und Anpassung der beiden Modellsätze aneinander wird anhand der Anpassungsqualität ( $\chi^2$ ) das beste Modell ermittelt.

### 4.6.2 Einzelpartikel-Cryo-Elektronenmikroskopie

Grundsätzlich werden zwei Arten von Elektronenmikroskopen verwendet, die *scanning* (SEM) und *transmission* (TEM) Elektronenmikroskopie. Bei der *scanning electron microscopy* (SEM) [193] werden grössere biologische Proben mit Metall beschichtet, das über die Leitfähigkeit der Oberfläche die Ermittlung derer Struktur ermöglicht. Bei der *transmission electron microscopy* (TEM) ermöglicht eine räumliche Gesamtaufnahme, da die beschleunigten Elektronen die Probe durchdringen (Abb. 24A)[194-195]. Ein auf das Präparat treffendes Elektron kann mit der Probe auf verschiedene Arten wechselwirken. Das Elektron erfährt durch das Präparat keine Änderung seiner Bewegung und Geschwindigkeit (ungestreutes Elektron), es wird gestreut ohne dass seine Energie geändert wird, dafür aber die Richtung (elastisch gestreutes Elektron) oder das Elektron verliert Energie und somit Geschwindigkeit, was zur kleinen Richtungsänderung im Magnetfeld der Linsen führt (unelastisch gestreutes Elektron).

Wenn das Elektron unelastisch gestreut wird, eine zufällige Menge Energie an die Probe abgegeben und seine Geschwindigkeit und Winkel geändert. So gestreute Elektronen erzeugen dann das Hintergrundrauschen, sofern sie auf den Leuchtschirm eintreffen. Ein von der Probe elastisch gestreutes Elektron behält seine Energie, aufgrund des Streuwinkels lassen sich im Nachhinein Aussagen über die Struktur der Probe treffen (Abb. 24B). Die Wahrscheinlichkeit, ob ein Elektron elastisch oder unelastisch gestreut wird, nimmt mit der Ordnungszahl des Streuzentrums zu. Daher werden häufig schwere Elemente als Kontrastmittel in der Elektronenmikroskopie benutzt. Der im TEM entstehende Kontrast ist abhängig von der Größe der abgebildeten Objektstrukturen. Proteine kleinen Durchmessers

(<10nm) zeichnen sich dadurch aus, dass die gestreute Elektronenwelle einen deutlich anderen Weg durch die Objektivlinse nimmt als die ungestreute Elektronenwelle, so dass die aufgrund unterschiedlicher Weglänge auftretende Phasendifferenz bei der Vereinigung beider Wellenzüge in der Nähe der Fokusebene zu Interferenzerscheinungen führt. Dies bezeichnet man als Phasenkontrast. Die Abbildung grössere Proteine (>10nm) wird nicht durch den Wellencharakter der Elektronen und den damit verbundenen Interferenzerscheinungen bestimmt. Die dabei entstehende Phasenverschiebung ist zu gering um ins Gewicht zu fallen. Sie werden im wesentlichen durch den Streuabsorptionskontrast abgebildet. Das Bild entsteht durch Ausblenden der unter großen Winkeln an schweren Elementen elastisch gestreuten Elektronen und den unabgelenkt passierenden Elektronen, die keine Wechselwirkung mit dem Präparat erfahren. So kommen auf den Leuchtschirm unterschiedlich viele Elektronen pro Fläche an und erhellen ihn unterschiedlich stark. Ein amplitudenmoduliertes Bild entsteht. Die entstandenen Bilder werden klassifiziert und nach dem entsprechenden Winkel in Projektionsklassen unterteilt. Anhand dieser kann aus dem zweidimensionalen Bilder ein dreidimensionales Modell erstellt werden (Abb. 24C).

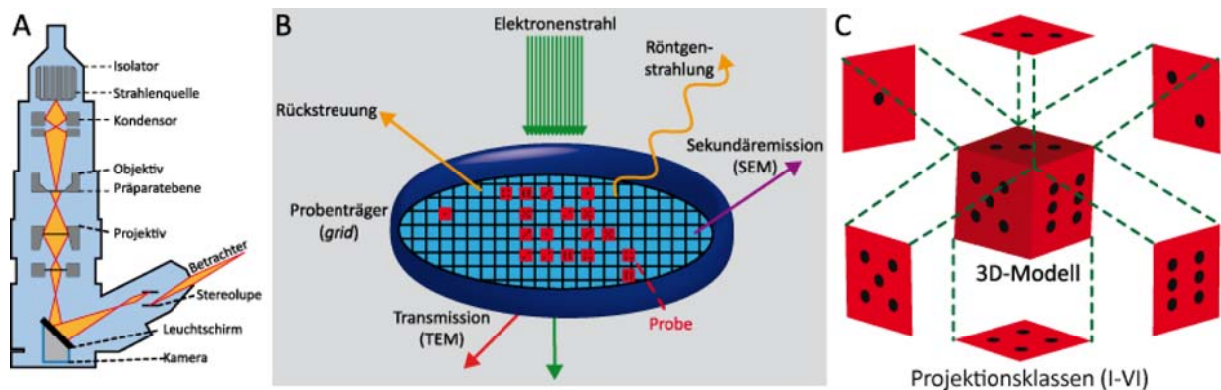


Abbildung 24: Strahlengang eines Elektronen-Mikroskops (A), gestreuter Elektronenstrahl am Probensträger (B) und iterative Modellrekonstruktion ausgehend aus den verschiedenen Projektionsklassen (C).

## i.) *Probenvorbereitung und Datensammlung*

Für biologische Proben wird auf wenige Millimeter grosse Kupferscheiben (*grids*) eine sehr dünne ionisierte Kohlenstoffschicht durch Graphitevaporation aufgedampft[165]. Diese sorgt für sehr starke ionische Interaktionen mit den Proteinmolekülen, so daß eine Elektronenmikroskop-Aufnahme erst möglich wird. Um den Kontrast einer Probe zu erhöhen, wird die

Proteinprobe mit einer Kontrastlösung (*stain*) versetzt. Die Kontrasterhöhung wird durch den Zusatz Schwermetalsalze (z.B. 2% Uranylacetat) erreicht. Dieses Verfahren wird *negative staining* genannt. Um die Probe während der Aufnahme vor der Strahlung zu schützen werden die mit der Probe beladenen *grids* in flüssigem Ethan ( $-188^{\circ}\text{C}$ ) eingefroren (*cryo electronmicroscopy*).

### ii.) Klassifizierung und Modellaufbau

Um aus zweidimensionalen Projektionen des Zielproteins ein dreidimensionales Bild zu rekonstruieren, werden mehrere Tausend experimentelle Projektionen benutzt. Für die Rekonstruktion eines Einzelpartikels (*single particle reconstruction*) ist es notwendig, diese vielen Projektionen in Klassen zu einzuordnen, dabei werden die Euler-Winkel jeder der Projektionen bestimmt. Die Euler-Winkel zeigen die Position und Orientierung der Projektion des Zielproteins an einer hypothetischen Sphäre, die das Zielprotein umspannt. Dabei definiert  $\theta$  die Höhe über dem Äquator,  $\phi$  die Rotation um den Äquator (Richtungswinkel oder Azimut) und  $\psi$  die Rotation um den von  $\theta$  und  $\phi$  definierten Zentrum der Sphäre. Um die Rekonstruktion des Zielproteins zu generieren, müssen zuerst alle Werte für  $\theta$  und  $\phi$  aller Projektionen ermittelt werden, um ihre Position auf der Euler-Sphäre zu bestimmen. Das Zentrum der Euler-Sphäre ist der Ursprung des dreidimensionalen Modells (*random conical tilt*). Der Prozess der Modellbestimmung geschieht iterativ. Die experimentell bestimmten Projektionen werden einer ersten Selektion unterzogen, anschließend werden sie zentriert, ausgerichtet und klassifiziert. Von so klassifizierten Projektionen wird ein erstes Modell erstellt (Abb. 24C). Um diese erste Lösung zu verfeinern, wird das Modell für eine bessere Ausrichtung und Klassifizierung der Projektionen benutzt. Aus den verfeinerten Daten werden neue Modelle erstellt, der Prozess wird so lange wiederholt, bis ein zufriedenstellendes Ergebnis erzielt worden ist.

### 4.6.3 Kristallographie

Auch wenn die Strukturaufklärung von Makromolekülen mit neuen Faltungsmotiven keinen exponentiellen Wachstum mehr aufweist, steuert die Strukturaufklärung einen wichtigen Beitrag bei der Beantwortung der biochemischen Schlüsselfragen[196]. Zur Untersuchung struktureller Eigenschaften von Makromolekülen können mehrere Methoden genutzt

werden, sofern aber Kristalle des Zielmoleküls vorliegen, wird die, wegen ihrer hohen Aussagekraft, röntgenkristallographische Untersuchung bevorzugt. Der Kristallisationsprozess wird in zwei Phasen vollzogen, der Keimbildung und dem Kristallwachstum. In der ersten Phase, der Keimbildung, müssen die Makromoleküle eine Energiebarriere überwinden, das Bilden eines periodisch geordneten Aggregatzustandes einer kritischen Größe, des Kristallisationskeimes, das in der Mutterlösung im thermodynamischen Sinn überleben kann[197]. Die zweite Phase, das Kristallwachstum, zeichnet sich über die höhere Attraktivität des festen Aggregates gegenüber dem flüssigen Zustand für jedes Makromolekül aus. Um in die Phase der Keimbildung und damit auch in die Kristallwachstumsphase zu gelangen, bringt man in einer Mutterlösung das ausgewählte Makromolekül in den Zustand der Übersättigung. Die Übersättigung ist ein Ungleichgewicht mit der Tendenz, das feste Aggregat zu vermehren bis der Zustand des Gleichgewichts wieder hergestellt ist. Ist in einer übersättigten Lösung die Keimbildungsbarriere überwunden, können sich Kristalle bilden und bis zu dem Einstellen des Gleichgewichts wachsen.

Bei der Kristallisation versucht man langsam in den Bereich der Übersättigung zu gelangen[198]. Dazu wird eine Proteinelösung mit Fällungsmitteln verschiedener Konzentrationen zu einer Mutterlösung kombiniert und zusätzlich durch den Einsatz an verschiedenen Puffer- und Salzlösungen die Löslichkeit des Proteins variiert. Im Idealfall befindet sich die Mutterlösung dann in einem Bereich nahe der Sättigungsgrenze. Die Bildung von Kristallen ist in dieser Phase noch nicht möglich, da in dem Untersättigungsbereich die Kristallbildung wegen der zu hohen Energiebarriere unmöglich ist, in der metastabilen Zone können aus bereits vorhandene Keimen Kristalle wachsen. Durch das Prinzip der Gasphasendiffusion in einer abgeschlossenen Kammer, in der sich ein Tropfen mit der Mutterlösung neben einem Reservoir befindet, wird mit der Zeit die Konzentration des Proteins in der Lösung langsam erhöht. Dadurch gelangt die Mutterlösung in den Übersättigungsbereich (Keimbildungszone). Dieser Vorgang ist von vielen Faktoren abhängig und lässt sich nicht voraussagen, diese besten Voraussetzungen erreicht man mit einer hohen Varianz der Mutterlösung sowie der Kristallisationstemperatur[199]. In der Keimbildungszone können sich aus dem spezifischem Aggregat Kristallisationskeime bilden. Durch die Bildung der Kristallisationskeime sinkt wiederum die Konzentration des Proteins in

der Mutterlösung womit die metastabile Zone erreicht wird, in der dann die Kristallisationskeime zu Proteinkristallen wachsen können. Wenn die Protein- bzw. Fällungsmittelkonzentration in der Mutterlösung zu hoch ist, wird die Zone der Bildung von Präzipitat in dem Übersättigungsbereich erreicht, in der das Protein sofort aggregiert und für die Kristallisation nicht mehr zu Verfügung steht.

## i.) Datensammlung und -prozessierung

Für die Datensammlung an der Drehanode oder einer Beamline (Tab. 12) werden die gezüchteten Kristalle an einem Cryo-Loop montiert und im flüssigem Stickstoff schockgefroren[200]. Der Kristallaufbau eines Proteinkristalls wird von den Proteinmolekülen bestimmt, die ein Gitter ausbilden, in den Hohlräumen dazwischen kann sich das Solvens einlagern (30% - 70%). Um die Kristalle vor Beschädigungen durch die schlagartige Ausdehnung des eingelagerten Solvens, oder Bildung störender Eiskristalle zu schützen werden sie vor dem Einfrieren in eine Cryo-Lösung überführt[201].

Modul	EMBL BW7A HasyLAB	ID14-2 ESRF Grenoble	X06SA SLS Villigen
Wellenlänge	0.7 - 1.8 Å	0.9333 Å	variabel
Detektor	MAR CCD 165 mm	ADSC Q4 CCD	PILATUS 6M

Tabelle 12: Instrumentspezifikation der in dieser Arbeit verwendeten Beamlines

Für die Montierung des Kristalls wird er am Mikroskop (Olympus) in einer Schlaufe eines Cryo-Loops (Schlaufengröße von 0.1 - 0.5 mm; *Hampton Research*) fixiert[202]. Die so schockgefrorenen Proteinkristalle wurden mit einer Cryo-Zange bzw. durch einen Probenwechsler auf den Goniometerkopf einer Beamline montiert. In einem Flüssigstickstoff-Strom werden die Kristalle während der Datensammlung bei 100 K gekühlt. Nach der Ausrichtung des Kristalls im Röntgenstrahl wird die Messung gestartet. Anhand einer Testaufnahme bei einem Drehwinkel von 0.3° lassen sich Mosaizität, Auflösung, Raumgruppe und Zellparameter für ein Proteinkristall bestimmen. Für die Prozessierung der gesammelten Daten wurden mehrere anwendungsorientierte Programme verwendet. Die Integration der gesammelten Daten erfolgte durch das Programm XDS[203] und XSCALE[204], die Skalierung der Intensitäten der Einzelbilder wurde mit dem Programm SCALA[205] durchgeführt.

*ii.) Strukturlösung und Verfeinerung*

Der Erhalt von weit streuenden Proteinkristallen und deren Vermessung sind nicht mit der Aufklärung der Proteinstruktur gleichzusetzen. Der Grund dafür liegt in der Art der Messung selbst, denn die Daten, die für die Strukturaufklärung benötigt werden, sind nicht alle aus der Messung selbst zu ermitteln. Der Proteinkristall wird während der Vermessung in einen monochromatischen, nicht kohärenten Synchrotronstrahl gehalten. Die Strahlung wird dabei an dem Kristallgitter gebeugt und die Intensität der gebeugten Strahlen an einem Detektor gemessen. Trifft der Strahl auf den Kristall, kommt es wegen der verschiedenen Netzebenenabstände zu miteinander interferierenden Röntgenstrahlen. Jedes Teilgitter des Kristalls beugt die einfallenden Röntgenstrahlen im selben Winkel. Die verschiedenen Ebenen (gleiche *hkl*-Indizes) führen zum Phasenunterschied der im selben Winkel gebeugten Welle[206]. Aufgrund der Beugung kommt es zu einer Interferenz und somit der Verschiebung der einzelnen Phasen untereinander. Konstruktive (gerade ganzzahlige Differenzen der Wellenlängen  $n \cdot 2\pi$ ) bzw. destruktive (gerade halbzahlige Differenzen der Wellenlängen,  $n \cdot \frac{1}{2}\pi$ ) Interferenz kann man aus erscheinenden Reflexen bzw. systematischen Auslöschungen erschließen. Aus diesen Auslöschungsregeln kann die Raumgruppe ermittelt werden, sowie weitere Aussagen über die Symmetrie, aber nicht die Lage der Atome in der Einheitszelle. Der Strukturfaktor  $F$  ist die Gesamtergebnis der in einer Richtung eines *hkl*-Reflexes (Millerschen Indizes) gebeugten Wellen. Die gemessene Intensität  $I$  ist proportional zum Quadrat vom Betrag des Strukturfaktors  $F$ , und der Strukturfaktor  $F$  ist die Fourier-Transformation der Elektronendichte  $\rho$ . Die gemessenen Intensitäten besitzen nur Informationen über die Amplitude, also den Betrag des Strukturfaktors. Um aus den Intensitäten den Strukturfaktor abzuleiten, muss deshalb das Phasenproblem gelöst werden. Für Proteinstrukturen können die Phasen bevorzugt durch mehrere Methoden gelöst werden (SAD/MAD, SIR/MIR, MIRAS, MR). Durch den Erhalt der Phasen kann die gemessene Elektronendichte berechnet werden in die ein Strukturmodell des Zielproteins angepasst wird. Dabei wird ein Teil der Strukturfaktoren bei dem Anpassungsprozess separiert ( $R_{\text{free}}$ ), während der übrige Teil ( $R_{\text{work}}$ ) durch die iterative Strukturmodell-Anpassung an die Elektronendichte (Verfeinerung) selbst angepasst wird. Während dem Prozess der Verfeinerung wird das Strukturmodell stetig an die gemessene Elektronendichte angepasst



und die  $R_{\text{work}}$ -Strukturfaktor-Werte sinken. Da die freien  $R_{\text{free}}$ -Strukturfaktoren während dieser Prozedur nicht verändert werden, stellen sie den Kontrollmechanismus für die Integrität des Strukturmodells dar. Entwickeln sie sich parallel zu den  $R_{\text{work}}$ -Strukturfaktor-Werten bestätigen sie die Integrität des Strukturmodells, andernfalls stimmt die ermittelte Elektronendichte und das erstellte Strukturmodell nicht überein. Nachdem das richtige Strukturmodell für die ermittelte Elektronendichte erstellt worden ist, können die Strukturdaten wie Atompositionen und die Elektronendichte in die PDB-Datenbank[207] geladen werden.

## ii.a) MAD – multiwavelength anomalous diffraction

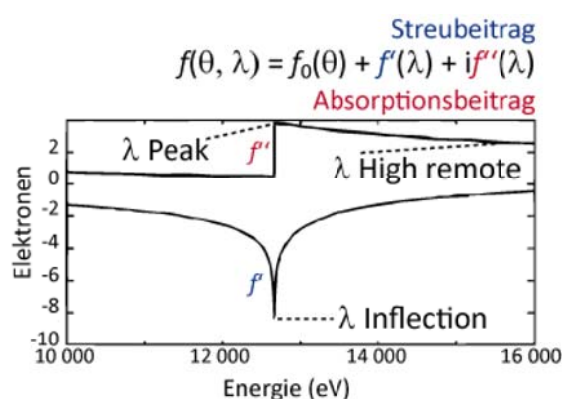


Abbildung 25: Selenabsorptionsspektrum

Bei der anomalen Dispersionsmethode (SAD/MAD)[208] absorbieren die inneren Elektronen eines Atoms Röntgenstrahlung einer bestimmten Wellenlänge und emittieren diese nach einer Verzögerung. In meisten Fällen wird die Dispersionsmethode an Selenatomen durchgeführt. Diese werden als Selenomethionin in das Zielprotein eingebaut, so dass ihre Position in der Sequenz fest zugeordnet

ist. Es werden drei Datensätze gesammelt, *inflection*-, *peak*- und *high remote*-Datensatz. Die so gemessenen Absorptionsspektren (Abb. 25) lassen unter Berücksichtigung des Friedelgesetzes Rückschlüsse über die Phasenverschiebung zu. Die Analyse dieser spezifischen Phasenverschiebung führt zu der Lösung des Phasenproblems.

## ii.b) MIR – multiple/single isomorphous replacement

Bei dem isomorphen Ersatz (SIR/MIR)[209] macht man sich das Prinzip zu nutze, dass die Diffusion des Kristallderivates (ein Kristall in das Schwermetallatome eingebracht sind) der Summe der Diffusionen des Ausgangskristalls und des hinzugefügten Schwermetallatoms beträgt. Da die Elektronen des schweren Atoms in einer relativ kleinen Sphäre konzentriert sind, ist ihre Diffusion in Phase, und ihr Beitrag produziert messbare Intensitätsänderungen im Beugungsmuster. Nur wenn das schwere Atom keine Veränderungen an der Proteinposition im Kristall verursacht, ist die Lösung des Phasenproblems auf diese Art

möglich. Wenn diese Methode benutzt werden soll, um eine nicht Zentrosymmetrische Kristallstruktur aufzuklären, muss mehr als ein isomorphes Derivat benutzt werden, in den meisten Fällen sind drei notwendig.

*ii.c) MIRAS – Kombination aus isomorphen Ersatz und anomaler Dispersion*

Bei dieser Methode nutzt man den Effekt, der entsteht, wenn die anomale Dispersion eines Atoms die gleiche wie bei dem schweren Atom ist, das durch den isomorphen Ersatz eingefügt wurde[210]. Die Methode des isomorphen Ersatzes führt zu einer Doppeldeutigkeit der Phase, da diese zu der Phase der anomalen Dispersion senkrecht ist, was zu einer Phasendetermination führt. Dies, macht man sich bei dieser Methode zur Lösung des Phasenproblems zunutze. Bei der MIRAS Methode werden mehrere schwere Atome benutzt, analog zu der MIR-Methode, die auch zur anomalen Dispersion benutzt werden können.

*ii.d) MR – molecular replacement*

Bei dem molekularem Ersatz (MR)[211-212] verwendet man die Daten von bereits bekannten Strukturen, die eine hohe Ähnlichkeit mit der erwarteten Struktur aufweisen (z.B bei homologen Proteinen, Mutanten, Protein-Inhibitor-Komplexen usw.). Durch Lösung der Rotations- und Translationsfunktion können die ähnlichen Bereiche als Basis für die Modellierung der neuen Elektronendichteverteilung dienen. Die Methode ist fehleranfälliger, da das Startmodell ohne experimentelle Phaseninformationen aufgebaut wird. Für diese Methode wurde das Program PHASER[213] benutzt.

## 4.7 Bioinformatische Methoden

Für die Auswertung der erstellten Alignments wurde eigens durch das *rapid prototyping* – Verfahren in der Scriptsprache Python ein Programm (für den Code siehe Anhang) entwickelt, um die Analyse von vielen Sequenzen bewerkstelligen zu können. Für die korrekte Ausführung werden zusätzlich Numerical- (*NumPy*) und Scientific- (*SciPy*) Python-Bibliotheken benötigt. Dabei wird das Alignment im *fasta*-Dateiformat als Input vorgegeben. In einer weiteren Kontrolldatei werden die Sequenzen, die als Begrenzungs-marker dienen, angegeben. Das Programm arbeitet dann die ausgerichteten Sequenzen der

Reihe nach ab, ermittelt die Anzahl der Aminosäuren zwischen den Begrenzungsmarkern ( $x_i$ ) und legt die Ergebnisse in eine Output-Datei ab. Diese ausgeschriebenen Daten bestehen aus einer Mittleren Länge über allen Sequenzen ( $\bar{x}$ ), deren Standardabweichung ( $\sigma$ ), der Verteilung der Aminosäuren für alle Sequenzen (0, siehe Anhang) sowie für jede einzelne Sequenz (von 1 bis N, siehe Anhang). Die Standardabweichung wird berechnet als die Quadratwurzel aus der Varianz, die die summierte quadrierte Abweichung vom Mittelwert aller Längen darstellt (Formel. 1).

$$\sigma_N = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}.$$

Formel 1: Standardabweichung

Um das für die Input-Dateien benötigte Alignment über alle Sequenzen zu erstellen, wurden die Sequenzen der verschiedenen NRPS-Systeme anhand entsprechender Kriterien in den Datenbanken Norine[80-81] und Pubmed gesammelt und anschliessend am *T-Coffee*[214]- bzw. *clustalW*[215]-Server ausgerichtet. Zusätzlich wurden Strukturmodell-unterstützte Verfahren am *3D-Coffee*[216]- und *Expresso*[217]-Server benutzt, um die Alignment-Untersuchungen an die Strukturdaten anzupassen. Alle Sequenzen mit einer erhöhten Abweichung (>50%) werden für spätere Betrachtungen nicht berücksichtigt.

## 5 Ergebnisse

Diese Arbeit beschäftigt sich mit der strukturellen Aufklärung nichtribosomaler Peptidsynthetasen. Wie bereits in der Aufgabenstellung dargestellt, wurde die Arretierung der dynamischen PCP-Domäne als Ausgangspunkt für diesen Ansatz gewählt. Um die noch ungeklärten Stationen der NRPS-Synthese untersuchen zu können, sind Informationen über die globale Struktur und Organisation der NRPS-Systeme notwendig. Um diese Aufgabe bewältigen zu können ist die Wahl verschiedener Zielproteine und Untersuchungsmethoden von großer Bedeutung. Der Fokus dieser Arbeit liegt in der Untersuchung unimodularer NRPS-Systeme mittels kristallographischer Methoden. Desweiteren steht die Analyse verschiedener Konformationen während der Naturstoffsynthese, die durch die Variation des PCP-Beladungszustands erreicht wird, sowie die Analyse der globalen Struktur multimodularer NRPS-Systeme durch Kleinwinkelstreuung (SAXS) und Cryo-Elektronenmikroskopie (CryoEM) ebenfalls im Mittelpunkt der Untersuchungen. Die Variation des PCP-Beladungszustands wurde nach einem tryptischen Verdau mittels massenchromatographischer Methoden verfolgt und erst bei einer einheitlichen Population weiter eingesetzt.

### 5.1 Kristallisation unimodularer NRPS

Für die strukturelle Bestimmung modularer NRPS-Systeme wurde, wegen der bereits positiven Voruntersuchungen, die S1003A-Variante von SrfAC aus *Bacillus subtilis* ATCC 21332 gewählt. Dafür wurde sie heterolog in *E. coli* BL21 (DE3) oder M15 Zellen exprimiert, über affinitätschromatographische Methoden aus dem Zelllysate extrahiert und anschließend durch das Gelfiltrations-Verfahren als monodisperse, definierte Proteinlösung gewonnen. Nach der Konzentrationsbestimmung wurde diese im *sitting drop*-Verfahren[218] unter verschiedenen, kommerziell erhältlichen Bedingungen und bei unterschiedlichen Konzentrationen auf ihre Kristallisationsfähigkeit getestet. Die so erstellten Kristallisationsplatten wurden bei 4° C oder 18° C inkubiert und in regelmäßigen Abständen kontrolliert. Die erhaltenen Ergebnisse wurden dokumentiert und unter der Verwendung der AXS FR591 Drehanode (*Bruker*) nach ihrem Streuverhalten klassifiziert. Als Proteinkristall identifizierte Kristalle wurden zwecks Datensammlung am Synchrotron untersucht. Zusätzlich wurde die Selenomethionin-

markierte SrfAC-Variante für die Phasenbestimmung (MAD-Verfahren) kristallisiert.

## 5.1.1 Heterologe Genexpression von SrfAC-S1003A

Für die Herstellung des rekombinanten Proteins SrfAC-S1003A im präparativen Maßstab wurde der *E. coli* BL21 (DE3) - pTrcHis2-SrfAC-S1003A Stamm im LB-Medium exprimiert. Zur Herstellung Selenomethionin-markierter Proteinkristalle wurde der SrfAC-S1003A tragende

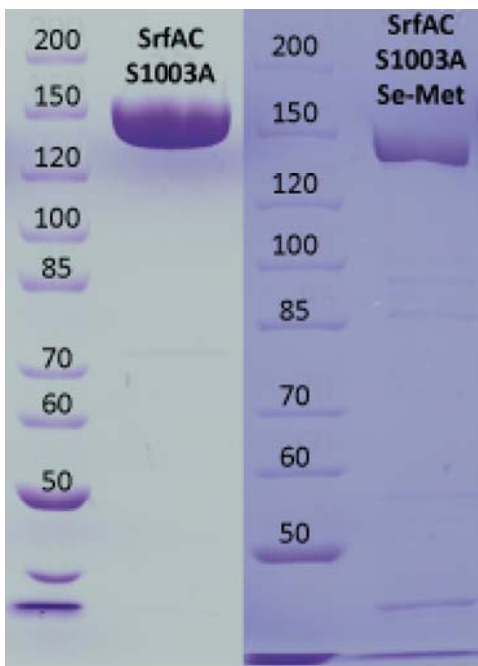


Abbildung 26: SDS-Gele des nativen und Selenomethionin-markierten SrfAC-S1003A.

Stamm im Minimalmedium M9 unter Zugabe von Selenomethionin exprimiert. Die Expression ergab eine Ausbeute unter nativen Bedingungen bis 20 mg Protein pro Liter Kultur, im Minimalmedium betrug die Ausbeute 5 mg pro Liter Kultur (Abb. 26). Die erste massenspektrometrischen Untersuchung der Inkorporation von Selenomethionin in dem heterolog produzierten SrfAC-S1003A ergab eine Besetzungszahl von bis zu 83%[219] für die verschiedenen Reste. Die Unterdrückung der Expression vor der Zugabe der Selenomethionin-Lösung durch die Addition von Glukose sollte die Besetzungszahl in dem optimalen Bereich halten. Durch die CD-Spektroskopie konnte die korrekte und

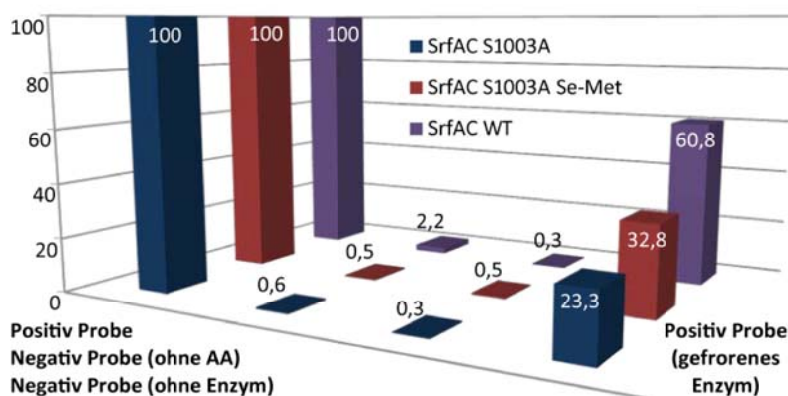


Abbildung 27: ATP/PPi-Austausch mit dem Wildtyp sowie der nativen und Selenomethionin-markierten SrfAC-S1003A-Variante mit Leucin als Substrat.

übereinstimmende Faltung des nativen SrfAC als auch der nativen und Selenomethionin-markierten SrfAC-S1003A-Variante bestätigt werden[219]. Die Aktivitätsuntersuchung der A-Domäne durch den ATP/PPi-Austausch zeigte, dass alle drei untersuchten Proteine (WT

sowie die native und Selenomethionin markierte S1003A-Variante) die korrespondierende Aminosäure Leucin aktivieren (Abb. 27). Da Kristallwachstum nur bei frisch hergestelltem Protein zu beobachten war, wurde ein Teil kurz vor dem Reaktionsstart im flüssigem Stickstoff eingefroren. Die eingefrorenen Proben zeigten eine deutliche Abnahme der Aktivität, die durch den Niedrigsalz-Puffer (10 mM HEPES pH 7.0, 100 mM NaCl) nur einen sehr geringen Cryo-Schutz erfahren. Für die Reproduktion und Optimierung der Kristalle wurden deshalb stets frisch hergestellte Proteinproben verwendet.

## 5.1.2 Kristallisation und Datensammlung von SrfAC S1003A

Die ersten Kristallisationsversuche wurden mit heterolog produziertem SrfAC-S1003A und einer Konzentration von 15 mg/mL, 15 mg/mL mit 5% Glycerol und 7.5 mg/mL durchgeführt. Die resultierenden Kristalle wurden ihrem Streuverhalten entsprechend klassifiziert (Tab. 13) [219]. Die Kristalle (Abb. 28) wurden unter anderem an der ESRF Beamline ID14-2 in Grenoble durch zwei um 90° versetzte Aufnahmen auf ihr Streuverhalten getestet (Projektbezeichnung MX659). Alle Kristalle, die proteintypisches Streuverhalten zeigten, wurden in Tabelle 13 aufgelistet. Die so erhaltenen Daten wurden automatisch einer Indizierung unterzogen, wobei die Raumgruppe, die Zellparameter sowie die theoretische

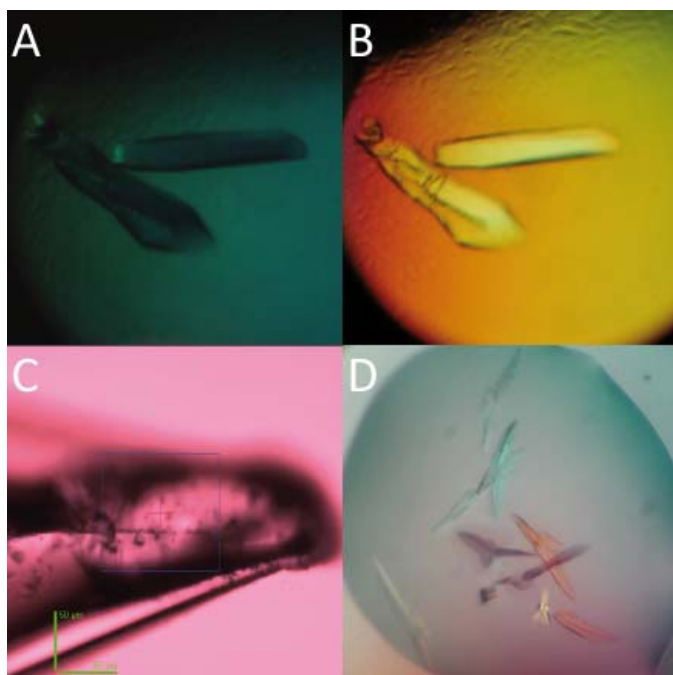


Abbildung 28: SrfAC-S1003 Kristallbilder. (A) Nativ (B) Nativ im Pol.-Licht (C) Montiert auf einem Cryo-Loop (D) Se-Met Kristalle.

Auflösung festgestellt wurden. Sofern die Kristalle eine zu geringe Auflösung ( $> 3.0 \text{ \AA}$ ), hohe Mosaizität ( $> 1.2^\circ$ ), zu schwache Streuintensität ( $I/\sigma < 2.0$ ) oder weiche (verschmierte) Reflexe zeigten, war eine automatische Indizierung nicht möglich (Tab. 13-k.l.). Das Kristallwachstum wurde vermehrt in den PEG 8000 und PEG 4000 enthaltenden Bedingungen festgestellt, wobei sich der pufferbedingte pH-Wert zwischen 5.5 und 8.5 bewegte. Auch eine Präferenz für

zweiwertige Kationen konnte festgestellt werden, wobei die  $\text{Ca}^{2+}$ -enthaltenen Bedingungen zu dendritisch-verzweigten, unregelmäßig gewachsenen Kristallen führten. Aus diesen Gründen wurde neben der Reproduktion der Kristalle auch die Optimierung der Kristallisationsbedingungen vorgenommen (Tab. 13).

Kristallisationsbedingungen für SrfAC S1003A [ $\text{Ca}^{\text{Leu}}\text{TTE}$ ] aus <i>Bacillus subtilis</i> ATCC 21332							
Nr.	Präzipitantz	Puffer	pH	Salz	t	Res.	Rep.
1	30% PEG 4000	0,1M TRIS	8,5	0,2 M $\text{LiSO}_4$	35	2.9 Å	k.l.
2	10% PEG 8000	0,1 M HEPES	7,5	-	5	DA 4.5 Å	3.0 Å
3	20% PEG 8000	0,1 M Na-Cacodylat	6,5	0,2 M $\text{Mg}(\text{OAc})_2$	11	k.l.	k.l.
4	10% PEG 8000	0,1 M Na-Cacodylat	6,5	0,2 M $\text{Mg}(\text{OAc})_2$	5	k.l.	k.l.
5	15% PEG 8000	0,1 M Na-Cacodylat	6,5	0,2 M NaOAc	28	k.l.	k.l.
6	8% PEG 10000	0,1 M Na-Citrat	5,6	0,1 M $\text{Mg}(\text{OAc})_2$	28	4.3 Å	k.l.
7	15% PEG 4000	0,1 M Na-Citrat	5,6	0,2 M $\text{NH}_4\text{OAc}$	28	k.l.	k.l.
8	15% PEG 4000	0,1 M TRIS	8,5	0,2 M $\text{MgCl}_2$	18	k.l.	k.l.
9	15% PEG 2000	0,05 M HEPES	7,5	0,1 M $\text{MgCl}_2$	35	DSn 2,6 Å	2,76 Å
10	15% PEG 4000	0,01 M TRIS	7,5	0,1 M LiCl	35	k.l.	k.l.
11	12% PEG 4000	0,1 M Na-Citrat	5,5	0,1 Mx( $\text{MgCl}_2$ +NaCl)	35	k.l.	k.l.
12	12% PEG 4000	0,1 M MES	6,5	0,1 Mx( $\text{MgCl}_2$ +NaCl)	18	k.l.	k.l.
13	12% PEG 4000	0,1 M HEPES	7,5	0,1 Mx( $\text{MgCl}_2$ +NaCl)	5	k.l.	DSs 2.8 Å
14	12% PEG 4000	0,1 M TRIS	8,5	0,1 Mx( $\text{MgCl}_2$ +NaCl)	18	k.l.	3.1 Å
15	10% PEG 8000	0,1 M MES	6,5	0,2 M $\text{Ca}(\text{OAc})_2$	18	k.l.	k.l.

Protein-Konzentration: a1: 15 mg/mL; a2: 15 mg/mL + 5% Glycerol; a3: 7 mg/mL;  
k.l. = keine Indexierung möglich; DA = Diplomarbeit; DSn = Datensatz nativ; DSs = Datensatz Se-Met

Tabelle 13: SrfAC S1003A Kristallisationswachstumsbedingungen.

Bei der Optimierung wurde die Konzentration der Präzipitantien zwischen 6 und 35% von PEG 2000, 4000, 8000 sowie 10000, mit wechselnden Pufferreagenzien wie TRIS, MES, HEPES, Citrat und Cacodylat und verschiedenen pH-Werten von 5.5 bis 8.5 variiert. Zusätzlich wurden parallel zu den Optimierungsbedingungen mit Puffer und Präzipans verschiedene  $\text{Mg}^{2+}$ -,  $\text{Cd}^{2+}$ -,  $\text{Mn}^{2+}$ - bzw.  $\text{Co}^{2+}$ -Salze hinzugesetzt. Das beste Streuverhalten zeigten die Kristalle aus den Bedingungen 1, 9, 13 und 14 (Tab. 13). Die Kristalle von nativem SrfAC-S1003A (15 mg/mL; Nr. 9 Tab. 13) wurden in Cryopuffer (5%(w/v) PEG 2000, 0.05 M HEPES pH 7.5, 0.1 M  $\text{MgCl}_2$ , 30% Glycerol) schockgefroren, anschließend wurde ein Datensatz aufgenommen (DSn, Tab. 13). Die Kristalle des Selenomethionin markierten SrfAC-S1003A zeigten unter veränderten Bedingungen (12 mg/mL; Nr. 13 Tab. 13) des Reproduktions- und Optimierungsscreens mit dem Adenylierungs-Additiv (3  $\mu\text{M}$  ATP,  $\text{Mg}^{2+}$  und Leucin),

schockgefroren im Cryopuffer (12% (w/v) PEG 4000, 0.1 M HEPES pH 7.5, 0.1 M MgCl<sub>2</sub>, 0.1 M NaCl, 30% Glycerol), die höchste Auflösung von 2.8 Å (DSs, Tab. 13). Der native Datensatz (DSn) wurde in Grenoble (**F**) an der ESRF Beamline ID14-2 mit einem Quantum CCD Detektor (ADSC) und die Selenomethionin-Datensätze in PSI-Villigen (**CH**) an der SLS Beamline X06SA mit einem PILATUS-6M Detektor (Dectris) aufgenommen.

### 5.1.3 Strukturlösung und Verfeinerung von SrfAC S1003A

Die gesammelten Daten wurden indiziert und mit XDS[203] und XSCALE[204] in der Raumgruppe  $P2_1$  mit den Zellparametern  $a = 83.11$ ,  $b = 106.56$ ,  $c = 92.40$  und  $\beta = 97.353^\circ$  integriert (Tab. 14).

Datenstatistiken – Datensammlung								
Datensatz (Wellenlänge in Å)		Zellparameter (Å)	Auflösung (Å)	Mehrfach (einmalig) gemessene Reflexe	$R_{\text{merge}}^*$	$I/\sigma(I)^{\dagger}$	$B_{\text{Wilson}} (\text{Å}^2)$	Vollständigkeit
nativ	(0.9330)	$a=83.11$ , $b=106.56$ , $c=92.40$ , $\beta=97.353^\circ$	82.5-2.60	98719 (45564)	3.8 (48.2)	12.9 (2.0)	79.2	0.931 (0.679)
SeMet $\lambda 1$	<b>[Peak]</b> (0.9794)	$a=83.36$ , $b=106.91$ , $c=92.74$ , $\beta=97.115^\circ$	46.0-2.80	267578 (39838)	3.6 (38.9)	24.2 (5.3)	105.3	0.999 (1.000)
SeMet $\lambda 2$	<b>[Inflection]</b> (0.9792)	$a=83.36$ , $b=106.91$ , $c=92.74$ , $\beta=97.115^\circ$	46.0-3.10	197105 (29430)	8.3 (57.8)	15.3 (3.5)	58.1	0.999 (1.000)
SeMet $\lambda 3$	<b>[High Remote]</b> (0.9717)	$a=83.36$ , $b=106.91$ , $c=92.74$ , $\beta=97.115^\circ$	46.0-2.95	229438 (66759)	5.6 (55.7)	18.4 (3.6)	86.5	0.853 (0.997)

\* $R_{\text{merge}} = ((\sum \sum |I_j(h) - \bar{I}(h)|) / (\sum \sum I_j(h))) \times 100$ ; Die Werte für  $I$  stehen für die Schale mit der höchsten Auflösung.  
 $^{\dagger}$  Durch das Programm TRUNCATE errechnet.

Datenstatistiken – Phasing										
Auflösung (Å)	9.40	6.80	5.60	4.90	4.40	4.00	3.70	3.50	3.30	3.10
Bewertungsfaktor – Figure of Merit (FOM)	0.64	0.62	0.59	0.53	0.45	0.40	0.36	0.30	0.27	0.24
Gesamt FOM	0.39	Gesamt FOM nach DM								0.80

Datenstatistiken – Verfeinerung	
Auflösungsgrenzen (Å)	82.5-2.60
Reflexe (work; test)	44479; 1040
$R$ -factor / $R_{\text{free}}^{\ddagger}$	0.213 (0.324) / 0.272 (0.436)
Aminosäuren	1274
Wasser Moleküle	39
Mittlere Wurzel aus der Summe der Bindungsabweichungsquadrate (RMSD bonds)	0.011 Å
Mittlere Wurzel aus Summe der Winkelabweichungsquadrate (RMSD angles)	1.272°

$\ddagger R = \sum |F_o| - k |F_c| / \sum |F_o|$ ; mit  $k$  als Skalierungsfaktor;  $R_{\text{free}}$  ist mit dem Test-Set berechnet.

Tabelle 14: Datenstatistiken: Datensammlung, Phasing und Verfeinerung.

Mittels des Matthews-Koeffizienten ( $2.87 \text{ Å}^3/\text{Da}$ ) wurde ein Solvensgehalt von 58.2% für ein Molekül pro ASU (*asymmetric unit*) berechnet. Für die Modellkonstruktion und Strukturlösung wurde das Programm PHASER[213] benutzt. Die Strukturdaten von PheA (pdb: 1AMU) für die A-Domäne sowie TycC<sub>6</sub>-C (pdb: 2JGP) für die C-Domäne wurden als Templat für die Erzeugung von Initialphasen nach der MR-Methode (*molecular replacement*)



verwendet. Durch die Initialphasen konnten alle 30 Se-Met-Reste in dem 2.8 Å MAD-Datensatz identifiziert und das erste SrfAC-S1003A-Modell, das 1179 der 1274 Aminosäuren beinhaltet, konstruiert werden (Tab. 14). Die resultierende Elektronendichte wurde durch das Programm DM[220] (*density modification*) strukturlösungsunabhängig optimiert. Das Strukturmodell wurde anschließend durch REFMAC5[221] und COOT[222] verfeinert, die Werte für  $R_{\text{work}}/R_{\text{free}}$  konvergierten bei 21.3%/27.2%. Die Polypeptidkette zeigt einen durchschnittlichen B-Faktor (Temperaturausdehnungskoeffizient) von 69.3 Å<sup>2</sup>, wobei der Wert in zwei Regionen erhöht (>90 Å<sup>2</sup>) war (Abb. 29A).

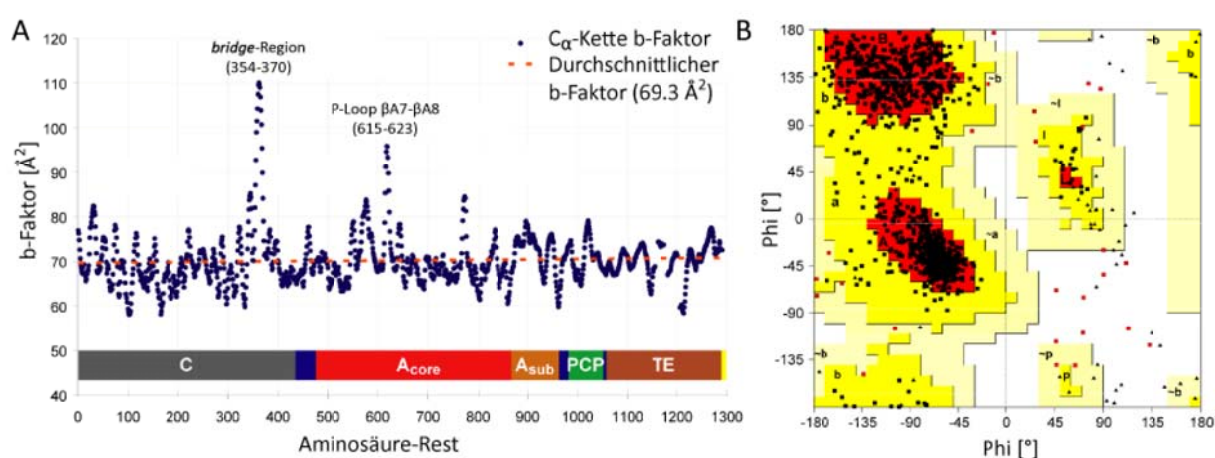


Abbildung 29: Diederwinkel und nach Aminosäure-Resten aufgelöste C<sub>α</sub>-B-Faktoren. (A) Die Polypeptidkette zeigt einen durchschnittlichen B-Faktor von 69.3 Å<sup>2</sup>. Für die bridge-Region (Aminosäuren 354-370) in der C-Domäne sowie für den P-Loop (Aminosäuren 615-623) in der A-Domäne können erhöhte (>90 Å<sup>2</sup>) B-Faktoren gemessen werden. (B) Ramachandran-Diagramm zeigt die Diederwinkel von 1128 von 1274 Resten (Ausgeschlossen Glycin- und Prolin-Reste), 961 Reste (85.2%) in den meistbevorzugten Region, 142 Reste (12.6%) in zusätzlichen erlaubten Region und 14 Reste (1.2%) im Randgebiet. Aus der schlecht aufgelösten Loop-Bereichen der Lid-Region aus der TE-Domäne befinden sich 11 Reste (1.0%) in nicht erlaubten Regionen. Die Analyse der Diederwinkel wurden mit dem Programm PROCHECK durchgeführt.

Den höchsten Wert zeigte die *Bridge-Region* der Kondensationsdomäne (βC11-αC11-αC12), vergleichbar mit TycC<sub>6</sub>[121] und VibH[117], gefolgt vom P-Loop der Adenylierungsdomäne (βA7-βA8), der in den Strukturen anderer Mitglieder der ANL-Familie in der offenen oder Adenylierungs-Konformationen sowie bei PheA[104], keine feststellbare Elektronendichte zeigt. Das Ramachandran-Diagramm (Abb. 29B) zeigt eine Verteilung der Diederwinkel von gesamt 1128 Resten (ohne Glycin- und Prolin-Reste), 961 Reste (85.2%) in der meistbevorzugten Region, 142 Reste (12.6%) in zusätzlich erlaubten Region und 14 Reste (1.2%) in dem Randgebiet. Nur 11 Reste (1.0%) lagen in der nichterlaubten Region, wobei diese in der sehr schlecht definierten Lid-Region der TE-Domäne lagen. Die Analyse der

Ramachandran-Diagramms wurde durch das Programm PROCHECK[223] durchgeführt.

## 5.2 Das SrfAC-S1003A-Strukturmodell

Kristalle wurden nur von der SrfAC-S1003A-Variante, bei der die, für die Bindung der prosthetischen Gruppe (PPant-Arm) verantwortliche Aminosäure Serin zu Alanin geändert wurde, erhalten. Ein vollständiger Datensatz wurde bei einer Auflösung von 2.60 Å aufgenommen und das Strukturmodell erstellt (Abb. 30).

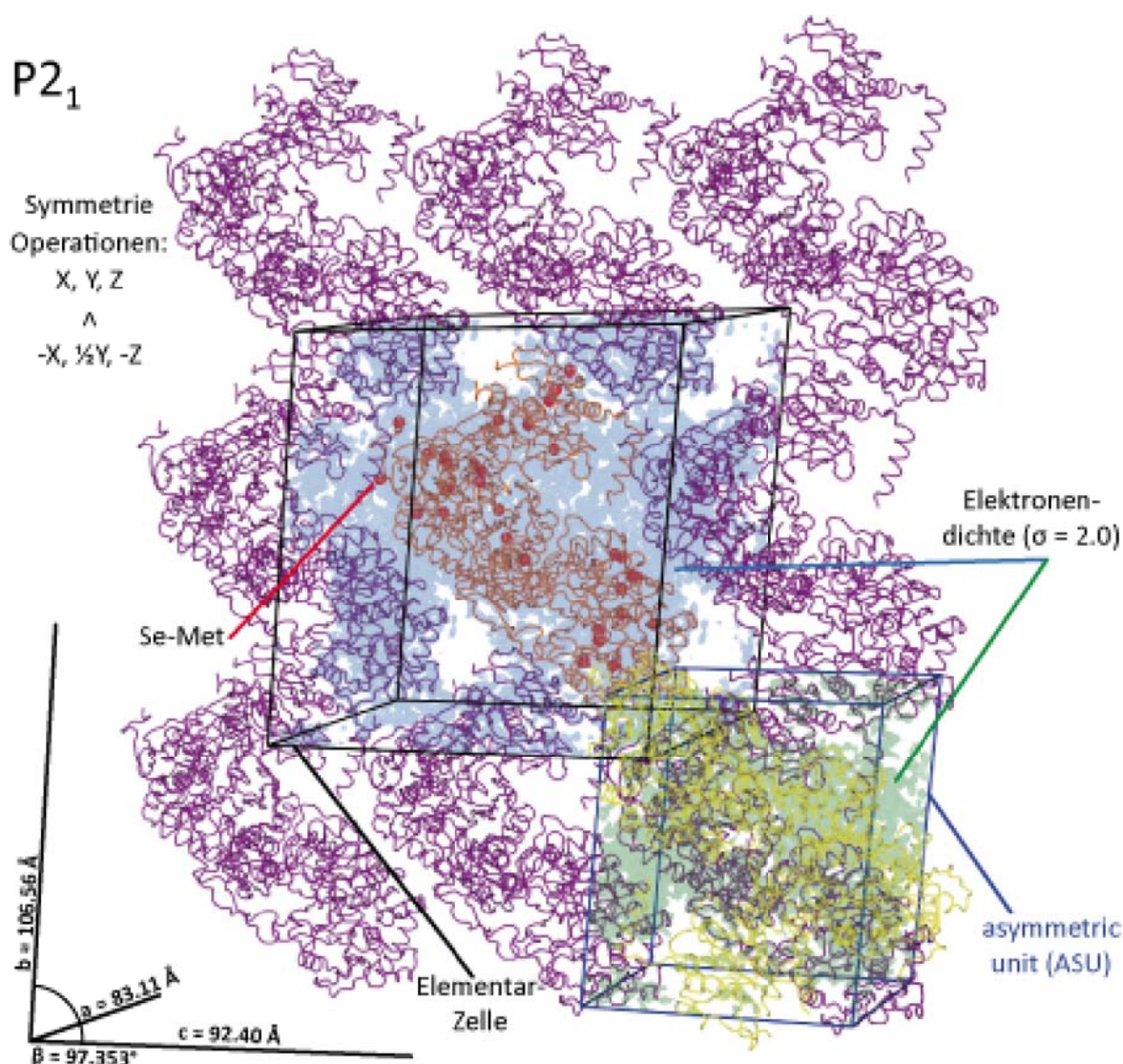


Abbildung 30: Kristallpackung für SrfAC-S1003A-Strukturmodell in der Raumgruppe P 2<sub>1</sub>. Dabei werden SrfAC-S1003A (orange) und Symmetrieäquivalenten (purpur) in der X-Z Ebene, sowie in ½Y-Richtung (gelb) dargestellt. Die Elektronendichte für die Elementarzelle (blau) sowie die ASU (grün) ist auf dem Signalevel von 2 angezeigt.

Das Strukturmodell beschreibt neben der SrfAC-S1003A-Sequenz von Methionin<sup>1</sup> bis Valin<sup>1274</sup> (ausgenommen mehrere ungeordneter *Loop*-Regionen mit undefinierter Elektronendichte) einen zusätzlichen C-terminalen Bereich (Threonin<sup>1269</sup> bis Glutaminsäure<sup>1290</sup>) der aus dem *myc*-Epitop sowie dem His<sub>6</sub>-Tag zusammengesetzt ist. Das Strukturmodell beschreibt 1179 von 1274 Aminosäuren, alle vier Domänen sind durch die Elektronendichte beschrieben.

### 5.2.1 Katalytische und dynamische Domänen

Das Terminationsmodul SrfAC des Surfactin Biosynthesecusters (Abb. 31A) aus *Bacillus subtilis* ATCC 21332 (144 kDa; vier Domänen [C-A-PCP-TE]; 1274 Aminosäuren) katalysiert sowohl den Einbau von L-Leucin in den Heptapeptid-Vorläufer als auch dessen Zyklisierung (Abb. 31B). Das Surfactin Biosynthesecuster ist in drei Synthetasen organisiert: die trimodulare Initiationsreaktion katalysierende Synthetase SrfAA[77], die trimodulare Elongationssynthetase SrfAB und die unimodulare terminale Synthetase SrfAC. Die globale Struktur kann in drei räumlich voneinander unabhängige Bereiche unterteilt werden: das Elongationsmodul, mit der interagierenden C-, A-, und PCP-Domäne, die TE-Domäne und die C-terminale *Tag*-Helix, bestehend aus *myc*-Epitop und His<sub>6</sub>-tag (Abb. 31C). Die für die Zyklisierung und die anschließende Freisetzung des  $\beta$ -Hydroxy-Fettsäure-Heptapeptids Surfactin verantwortliche TE-Domäne zeigt eine nahezu identische Faltung im Vergleich zu der alleinstehenden TE-Domäne (pdb: 1JMK)[124] mit einem RMSD-Wert (*root mean square deviation*) von 0.72 Å für 201 C $\alpha$ -Atome. Neben der für das Terminationsmodul essentiellen TE-Domäne bildet das Strukturmodell, repräsentativ dargestellt durch die Domänenabfolge C-A-PCP, das Rückgrat eines Elongationsmoduls. Das Elongationsmodul ist in den Typ A (linear) und Typ B (iterativ) NRPS-Systemen für die Aktivierung und den Einbau korrespondierender Aminosäure in den wachsenden Peptidvorläufer verantwortlich. Das Hauptgerüst dieser katalytischen Plattform bilden die Kondensationsdomäne und der Kern der Adenylierungsdomäne (A<sub>core</sub>), die zusammen 83% der Gesamtmasse ausmachen (118 kDa). Auf der Oberfläche dieser katalytischen Plattform agieren zwei kleinere Domänen, die 80 Aminosäuren große PCP-Domäne (Asparagin<sup>971</sup> – Glycin<sup>1041</sup>) und die 94 Aminosäuren große A<sub>sub</sub>-Domäne (Asparaginsäure<sup>861</sup> – Prolin<sup>955</sup>).



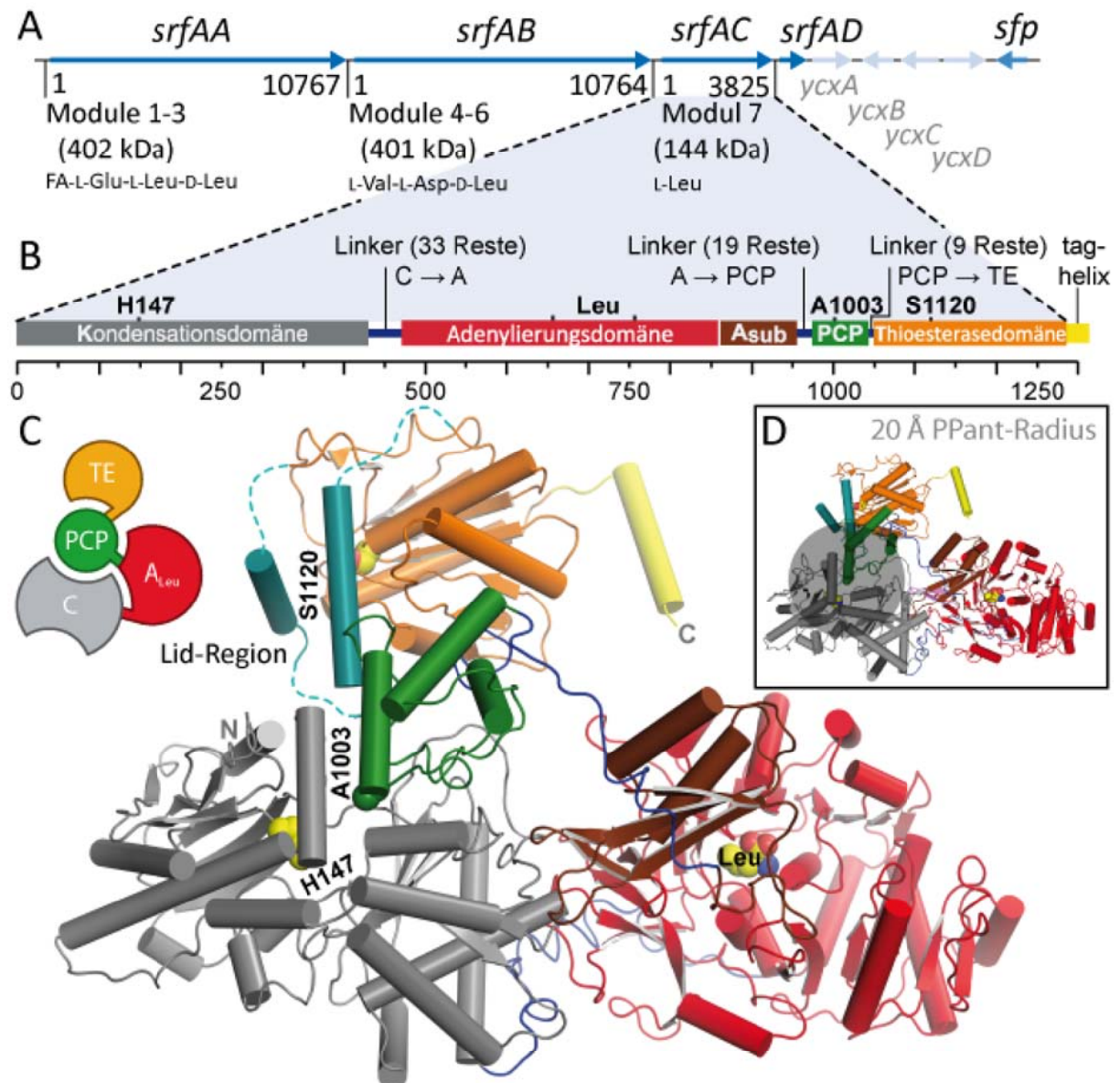


Abbildung 31: Organisation und Struktur der SrfAC-Synthetase aus *B. subtilis* ATCC 21332. (A) Organisation des Surfactin Biosynthesegenclusters in Synthetasegenen *srfAA*, *srfAB* und *srfAC*, externe Thioesterase-TypII-Gen *srfAD*, sowie das Phosphopantothetransferasegen *sfp*. Die drei Surfactin Synthetasen: Die trimodulare Initiationsynthetase SrfAA (402 kDa) katalysiert die Initiationsreaktion mit der C<sub>13-15</sub>-β-Hydroxy-Fettsäure, die um L-Glu, L-Leu und D-Leu elongiert wird, die trimodulare Elongationssynthetase SrfAB (401 kDa) katalysiert die Elongation des Vorläufers um L-Val, L-Asp und D-Leu und die unimodulare Terminationsynthetase SrfAC (144 kDa), die die finale Elongation durch L-Leu sowie die Freisetzung des zyklisierten Produktes katalysiert. (B) Domänenorganisation der SrfAC-S1003A-Variante. Die aktiven Zentren der jeweiligen Domänen sind hervorgehoben: Histidin<sup>147</sup> in der C-Domäne, das in der A-Domäne komplexierte Leucin (verantwortliches Bindungsmotiv ist zwischen den A4- und A5-Core-Motiven lokalisiert), die Alanin<sup>1003</sup>-Variante steht für das aktive Serin der PCP-Domäne und das aktive Serin<sup>1120</sup>, das als Teil der katalytischen Triade für die zyklische Freisetzung von Surfactin verantwortlich ist. (C) Strukturmodell von SrfAC-S1003A, als schematische und *Cartoon*-Strukturmodell Darstellung. Die katalytischen und dynamischen Kompartimente sind unterschiedlich eingefärbt: Die katalytischen Kompartimente wie die C-Domäne, A-Domäne und TE-Domäne sind in grau, rot und orange und die dynamischen Bereiche wie PCP-Domäne, A<sub>sub</sub>-Domäne und die flexible Lid-Region der TE-Domäne in grün, braun und petrol eingefärbt. Die Linker-Bereiche sind in blau dargestellt. (D) Wirkungsbereich der prosthetischen Gruppe von 20 Å mit Alanin<sup>1003</sup> als Ankerpunkt.

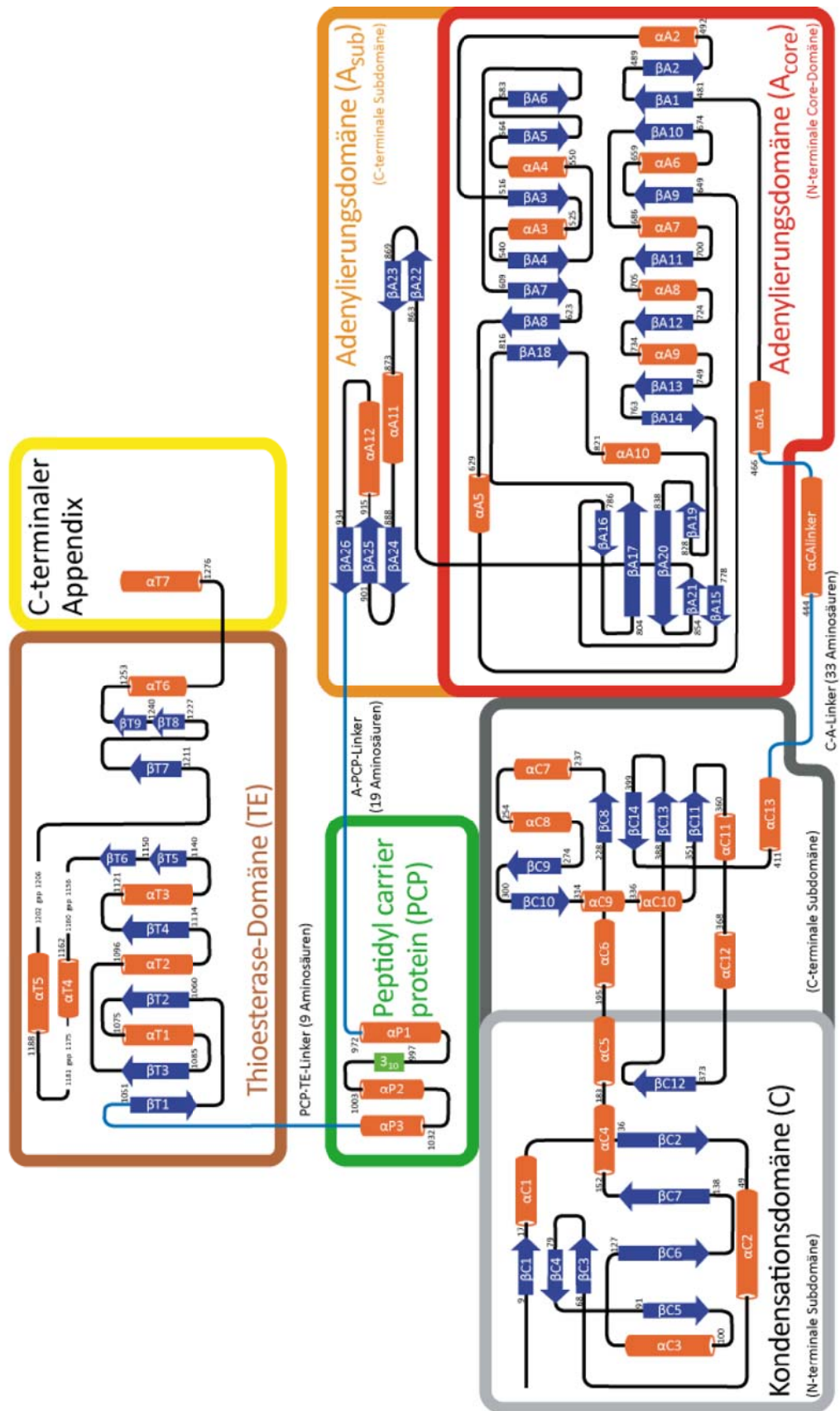


Abbildung 32: Sekundärstruktur Repräsentation des unimodularen SrfAC-Strukturmodells.

Die aktiven Zentren der C- und der A-Domäne sind 63 Å voneinander entfernt, markiert durch das Histidin<sup>147</sup> (C3-Motiv) und das durch die A-Domäne gebundene Leucin. Die das Alanin<sup>1003</sup> umspannende Sphäre, mit einem 20 Å langen Radius, steht für die theoretische Reichweite der prosthetischen Gruppe (Abb. 31D). Die große Distanz zwischen den einzelnen katalytisch relevanten Zentren kann die dynamisch agierende PCP-Domäne jedoch nur nach großen Konformationsänderungen überbrücken. Die katalytische Plattform, bestehend aus der starren C- und A<sub>core</sub>-Domäne, auf denen die dynamischen PCP- und A<sub>sub</sub>-Domänen agieren, wurde als *Workbench* definiert.

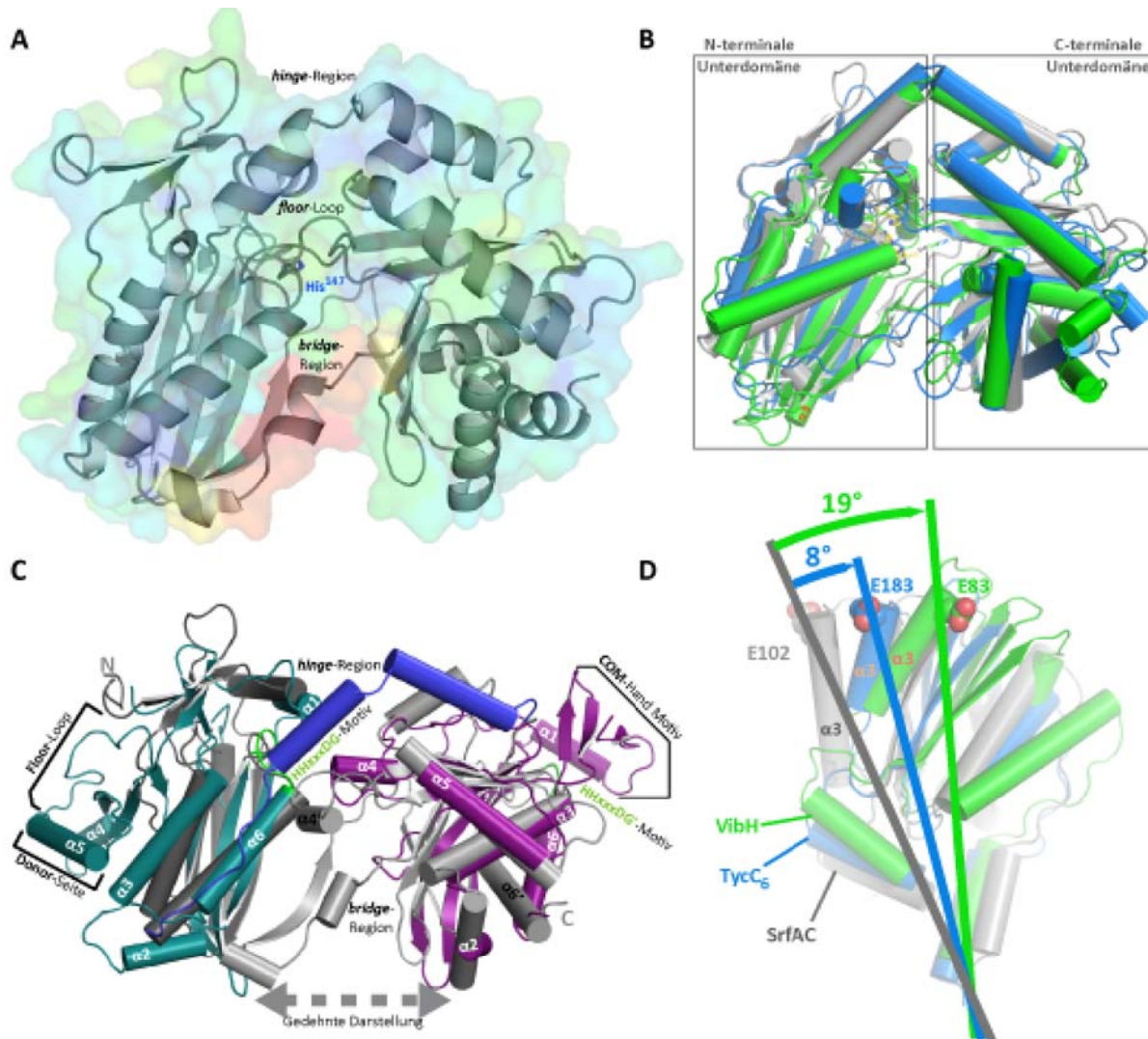
Die Kennzeichnung der einzelnen Domänen in den Abbildungen erfolgt nach dem NRPS-Code, bei dem die C-Domäne in grau, die A<sub>core</sub>-Domäne in rot, die A<sub>sub</sub>-Domäne in braun, die PCP-Domäne in grün und die TE-Domäne in orange, sowie alle Linker-Bereiche in blau eingefärbt sind. Die Bezeichnung der Sekundärstruktur-Elemente erfolgt chronologisch nach Domänen, wobei αC1 für die erste α-Helix in der C-Domäne steht (Abb. 32).

## i.) Kondensationsdomäne

Die C-Domäne (Methionin<sup>1</sup> – Asparagin<sup>431</sup>) zeigt eine pseudodimere Faltung[117] durch die Interaktion der N-terminalen (Methionin<sup>1</sup> – Glutamin<sup>193</sup>) und der C-terminalen (Asparaginsäure<sup>194</sup> – Asparagin<sup>431</sup>) Subdomäne. Beide Subdomänen zeigen eine CAT ähnliche Faltung (Chloramphenicol-Acetyltransferase; pdb: 3CLA) und sind über ein flexibles Gelenk (*hinge*) am Lysin<sup>192</sup> in einer V-Form zueinander angeordnet. Zwischen den Subdomänen sind drei Regionen definiert: *Hinge*- und *Bridge*-Region sowie der *Floor*-Loop (Abb. 33A). Während die *Hinge*-Region die Grenze zwischen den beiden Subdomänen definiert, bilden die *Bridge*-Region und der *Floor*-Loop die Reaktionskammer um Histidin<sup>147</sup>. Dabei zeigt die *Bridge*-Region die höchsten b-Faktoren, nicht nur in der C-Domäne, sondern in der gesamten SrfAC-Struktur (Abb. 33A). Als erstes Struktumodell einer <sup>D</sup>C<sub>L</sub>-Domäne, die die Kondensationsreaktion zwischen einer <sup>D</sup>-Aminosäure von der Donor-Seite mit der <sup>L</sup>-Aminosäure von der Akzeptor-Seite katalysiert zeigt die SrfAC <sup>D</sup>C<sub>L</sub>-Domäne die gleiche Faltung wie die TycC<sub>6</sub>-<sup>D</sup>C<sub>L</sub>-Domäne oder die alleinstehende VibH, die die Kondensationsreaktion zwischen PCP-gebundenem DHB und Norspermidin (NSPD) katalysiert (Abb. 33B). Die N-terminale Unterdomäne teilt das COM-Hand Motiv samt der αC1-Helix, die an der Interaktion mit der PCP verantwortlich ist, sowie den katalytisch relevanten Loop mit dem HHxxxDG-Motiv (β7-



$\alpha$ 4-Loop, bzw.  $\beta$ 7- $\alpha$ 6-Loop bei CAT) mit dem CAT-Monomer (Abb. 33C, cyan). Die C-terminale Untereinheit dagegen zeigt strukturelle Homologien zu dem Floor-Loop sowie der  $\alpha$ 5-Helix (Donor-Seite), die C-terminale  $\alpha$ 6-Helix nimmt jedoch eine abweichende Orientierung an (Abb. 33C, magenta).



**Abbildung 33:** Die SrfAC Kondensationsdomäne im Vergleich mit NRPS- und CAT-Vertretern. **(A)** SrfAC C-Domäne als Cartoon-Darstellung, die transparente Oberflächenrepräsentation ist den B-Faktoren entsprechend eingefärbt, wobei der rote Bereich sehr hohen ( $> 90 \text{ \AA}^2$ ) und der blaue niedrigen ( $< 60 \text{ \AA}^2$ ) B-Faktoren entspricht. **(B)** Struktureller Vergleich der C-Domänen Strukturmoleküle SrfAC-C (grau), TycC<sub>6</sub>-C (blau) und VibH (grün) über die Superposition der C-terminalen Subdomäne. **(C)** Die Superposition des N- und C-terminalen Subdomänen mit CAT-Monomer, zur besseren Übersicht sind die Subdomänen in einer räumlich gedehnten Repräsentation dargestellt. Während die N-terminale Subdomäne die COM-Hand und das HHxxxDG-Motiv mit dem CAT-Monomer teilen, zeigt die C-Terminale Subdomäne Übereinstimmungen mit dem Floor-Loop und der  $\alpha$ 5-Helix auf der Donor-Seite. Die Bridge-Region der C-Domäne weicht in ihrer Orientierung und Faltung stark von dem CAT-Monomer ab und ist zusammen mit der Hinge-Region den NRPS C-Domänen vorbehalten. **(D)** Seitenansicht der im absoluten Winkel voneinander abweichenden N-terminalen Subdomänen der C-Domänen Superposition.

Analog zu Sfp (Pseudodimer) und AcpS (Trimer) bildet das CAT-Trimer (Abb. 33A) die katalytisch aktive Furche zwischen den Monomeren aus, während bei der Kondensationsdomäne die Furche zwischen den zwei als Pseudodimer angeordneten Untereinheiten entsteht. Während der *Floor*-Loop auch in den verwandten CAT-Systeme strukturell konserviert ist, zeigen sich im Vergleich mit der *Bridge*-Region starke Unterschiede. Im Vergleich zu der Chloramphenicol- Acetyltransferase zeigt die Kondensationsdomäne auch einzigartige Strukturelemente wie die *Hinge*-Region (Abb. 33C blau), die zwischen dem N- und C-terminalem Bereich lokalisiert ist und die *Bridge*-Region, die den Boden der Reaktionsfurche definiert. Im Strukturvergleich mit anderen C-Domänen wie VibH (pdb: 1L5A)[117] und TycC<sub>6</sub> (pdb: 2JGP)[121] sind diese einzigartigen Bereiche in allen drei Strukturen definiert (Abb. 33B). Auch wenn die Faltung der C-Domänen identisch scheint, zeigen die zwei Subdomänen eine veränderte Neigung in Relation zu einander. Zu der C-Domäne aus SrfAC (pdb: 2VSQ) als Bezugspunkt zeigt die TycC<sub>6</sub> C-Domäne eine Neigung von 8° und VibH eine Neigung von 19° (Abb. 33D). Diese abweichende Neigung kann im zu bindenden Substrat oder in Domäneninteraktionen begründet sein. Die Faltung der C-Domäne ermöglicht den Zugang der PPant-gebundenen Substrate durch die Akzeptor- bzw. Donor-Seite, wobei im SrfAC-Strukturmodell nur die Akzeptor-Seite durch die PCP-Domäne besetzt ist.

## ii.) Adenylierungsdomäne

Auch wenn die zu der ANL-Familie zugehörigen NRPS-A-Domänen die artgleiche Reaktion katalysieren wie die tRNA-Synthetasen, wobei zuerst die korrespondierende Aminosäure aktiviert und anschließend auf die entsprechende Carrier-Domäne übertragen wird, zeigen sie weder evolutionäre noch strukturelle Verwandtschaft[224]. Alle Mitglieder der ANL-Familie teilen eine charakteristische Faltung. Diese kann in zwei Subdomänen, die A<sub>core</sub>- und A<sub>sub</sub>-Domäne unterteilt werden, wobei die A<sub>sub</sub>-Domäne während der Adenylierungsreaktion ihre Orientierung in Bezug zu der A<sub>core</sub>-Domäne wechselt. Die A<sub>core</sub>-Domäne ist durch das Zusammenspiel der drei Regionen A, B und C definiert[108], dabei zeigen die Regionen A und B eine analoge Faltung. Während die durch die drei Core-Motive A1, A2 und A3 definierte Region A den Threonin und Serin reichen P-Loop beinhaltet, ist die durch die Core-Motive A4 und A5 definierte Region B für die Erkennung und Bindung der korrespondierenden Aminosäure verantwortlich. Die Core-Motive A6 und A7 bilden ein verzerrtes  $\beta$ -Barrel, deren Loop-



Regionen an der Interaktion mit der C-Domäne beteiligt sind. Die A<sub>sub</sub>-Domäne kennzeichnen folgende drei Core-Motive: Das Core-Motiv A8 bildet das Scharnier zwischen A<sub>core</sub>- und A<sub>sub</sub>-Domäne sowie die  $\beta$ -Schleife, die während der Thiolierung den Pantotheintunnel formt, das Core-Motiv A9 bildet die  $\alpha$ -Helix für die PCP-Interaktion und das Core-Motiv A10 enthält das während der Adenylierungsreaktion relevante Lysin<sup>948</sup> (Abb. 34).

Die Elektronendichte zeigt ein assoziiertes Leucin in der A<sub>core</sub>-Domäne, das durch die für die

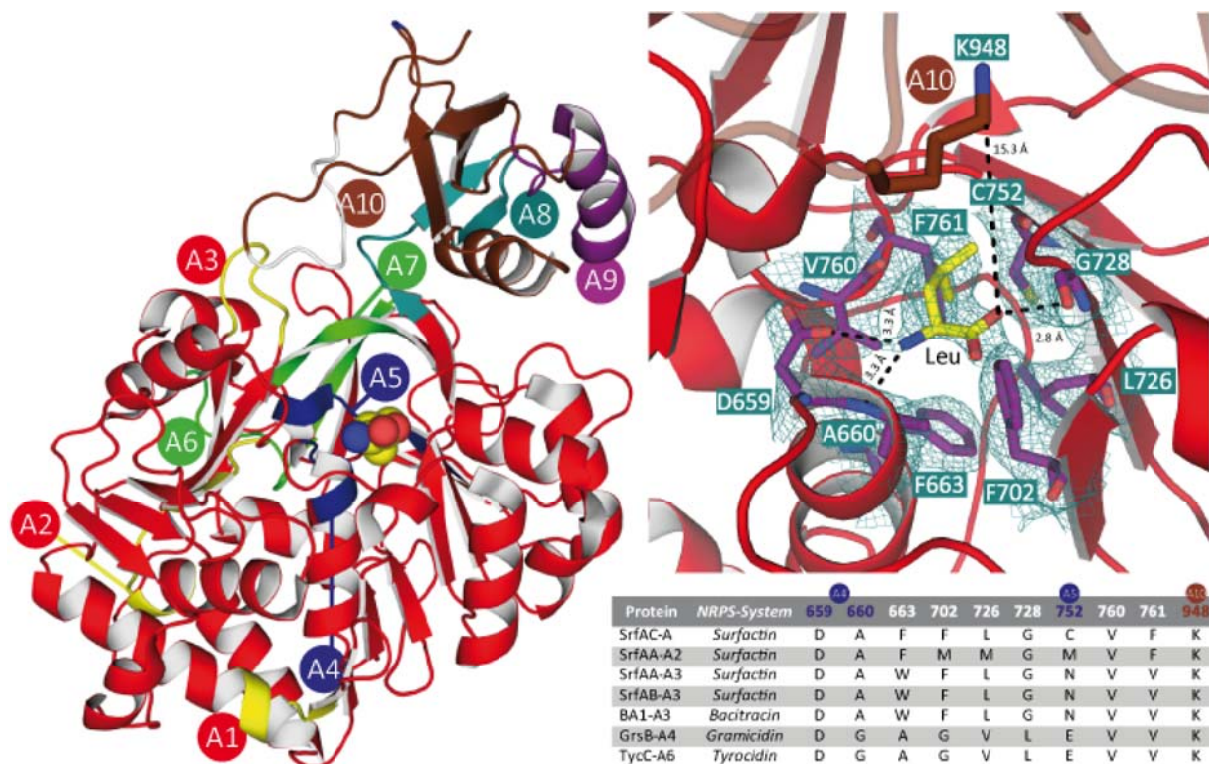


Abbildung 34: Die Leucin aktivierende SrfAC-A<sub>core</sub>- und A<sub>sub</sub>-Domäne und das Bindungsmotiv. Die SrfAC A-Domäne zeigt als Mitglied der ANL-Familie die charakteristische Faltung, die A<sub>sub</sub>-Domäne (braun) nimmt jedoch in Bezug zu der A<sub>core</sub>-Domäne (rot) eine neue Orientierung an. Die Core-Motive A1 und A2 (gelb) sind konservierte Strukturelemente in allen Mitgliedern der ANL-Familie. Der Threonin- und Serin-reiche P-Loop, der zu dem A3 Core-Motiv gehört, ist in der Elektronendichte definiert, zeigt aber nach der Bridge-Region der C-Domäne die höchsten B-Faktor-Werte. Die für die Substratbindung verantwortlichen Reste D<sup>659</sup> und A<sup>660</sup> (Distanz von jeweils 3.3 Å zum Leu-NH<sub>2</sub>) gehören zu dem Core-Motiv A4, der C<sup>752</sup>-Rest zum A5-Motiv (blau), der dazwischen liegende Bereich (F<sup>663</sup>, F<sup>702</sup>, L<sup>726</sup>) bildet zusammen mit den Resten V<sup>760</sup> und F<sup>761</sup> die hydrophobe Tasche, in der die unpolare Leucin-Seitenkette gebunden wird. Der Carboxylrest des gebundenen Leucins interagiert mit dem G<sup>728</sup>-Rest (2.8 Å). Neben den für die Substratbindung verantwortlichen Resten (purpur) konnte in der Elektronendichte ( $\sigma = 1.0$ ) auch das gebundene Leucin identifiziert werden. Die Core-Motive A6 und A7 (grün) bilden ein verzerrtes  $\beta$ -Barrel über den die Interaktion mit dem C-terminalem Bereich der C-Domäne erfolgt. Der flexible Bereich (Core-Motiv A8, grau) zwischen der A<sub>core</sub>- und A<sub>sub</sub>-Domäne ermöglicht die Konformationsänderung der Subdomäne. Die durch den Core-Motiv A9 (purpur) initiierte  $\alpha$ -Helix und das Core-Motiv A10 (braun) befinden sich nicht in der Reichweite der katalytischen Zentrum, das katalytisch relevante Lysin<sup>948</sup> aus Core-Motiv A10 befindet sich über 15 Å weit entfernt vom gebundenen Leucin. Die für die Substratbindung verantwortlichen Aminosäuren in den Surfactin-Cluster zeigen Unterschiede zu Gramicidin oder Tyrocin, folgen aber intern einem analogen Muster.

Substratbindung verantwortlichen Reste komplexiert wird. Der zwischen den Core-Motiven A4 und A5 liegende Bereich, der für die spezifische Substratbindung kodiert, zeigt im Vergleich zu den Leucin aktivierenden A-Domänen der Surfactin-Synthetasen keine vollständige Übereinstimmung, jedoch bestehen alle Reste aus unpolaren Aminosäuren (Abb. 34). Die Aminogruppe des gebundenen Leucins wechselwirkt mit der 3.3 Å entfernten Seitenkette des hochkonservierten Asparaginsäure<sup>659</sup> und mit der ebenfalls 3.3 Å entfernten Aminogruppe des Alanin<sup>660</sup>, beide Reste gehören zu dem Core-Motiv A4. Die Carboxylgruppe des gebundenen Leucins interagiert mit der 2.8 Å entfernten Carboxylgruppe des Glycin<sup>728</sup>, das während der Adenylierung wichtige Lysin<sup>948</sup> ist über 15 Å vom gebundenen Leucin entfernt. Die unpolare Seitenkette des Leucins wird durch das Substratbindungsmotiv (F<sup>663</sup>, F<sup>702</sup>, L<sup>726</sup>, C<sup>752</sup>, V<sup>760</sup> und F<sup>761</sup>; Tabelle Abb. 34) komplexiert. Dabei bilden die unpolare Seitenketten der entsprechenden Aminosäuren eine hydrophobe Tasche, aus in der die Leucin-Seitenkette lokalisiert ist (Abb. 34).

### iii.) *Peptidyl-Carrier-Protein (PCP) und die Thioesterase*

Bei der SrfAC-S1003-Variante fehlt der PCP-Domäne die Fähigkeit, die prosthetische Gruppe (PPant-Arm) zu tragen und ist daher in der AH-Konformation arretiert[111]. In der AH-Konformation zeigt die PCP-Domäne ihre intramodularen Interaktionen[121, 125]. Kristallwachstum mit dem Wildtyp in der Apo- oder Holo-Form, mit Thio-, bzw. Amino-CoA sowie deren Derivaten beladen, konnte nicht beobachtet werden. Die C-terminale TE-Domäne zeigt für mehrere *Loop*-Bereiche der *Lid*-Region (Leucin<sup>1156</sup> bis Threonin<sup>1160</sup>, Asparagin<sup>1175</sup> bis Glutamat<sup>1181</sup> und Isoleucin<sup>1202</sup> bis Glutamin<sup>1206</sup>) keine Elektronendichte (Abb. 35). Im Vergleich zu der offenen Konformation der alleinstehenden SrfAC-TE-Domäne (pdb: 1JMK-O) zeigt die TE-Domäne (pdb: 2VSQ) einen RMSD-Wert von 0.72 Å für 201 C<sub>α</sub>-Positionen, zu der geschlossenen Konformation (pdb: 1JMK-C) ein RMSD-Wert von 1.20 Å für 195 C<sub>α</sub>-Positionen[225]. Dabei weicht die Position der in der *Lid*-Region lokalisierten αTE4-Helix in der offenen Konformation von der in der geschlossenen Konformation ab. Die katalytische Triade, bestehend aus Serin<sup>1120</sup>, Asparaginsäure<sup>1147</sup> und Histidin<sup>1247</sup> (Serin<sup>80</sup>, Asparaginsäure<sup>107</sup> und Histidin<sup>207</sup> in 1JMK), dagegen kann übereinstimmend superpositioniert werden. Eine weitere Besonderheit ist die C-terminale *myc*-Helix, die entlang der X,Z-Kristallebene für die intermolekularen Kristallkontakte verantwortlich ist (Abb. 30).

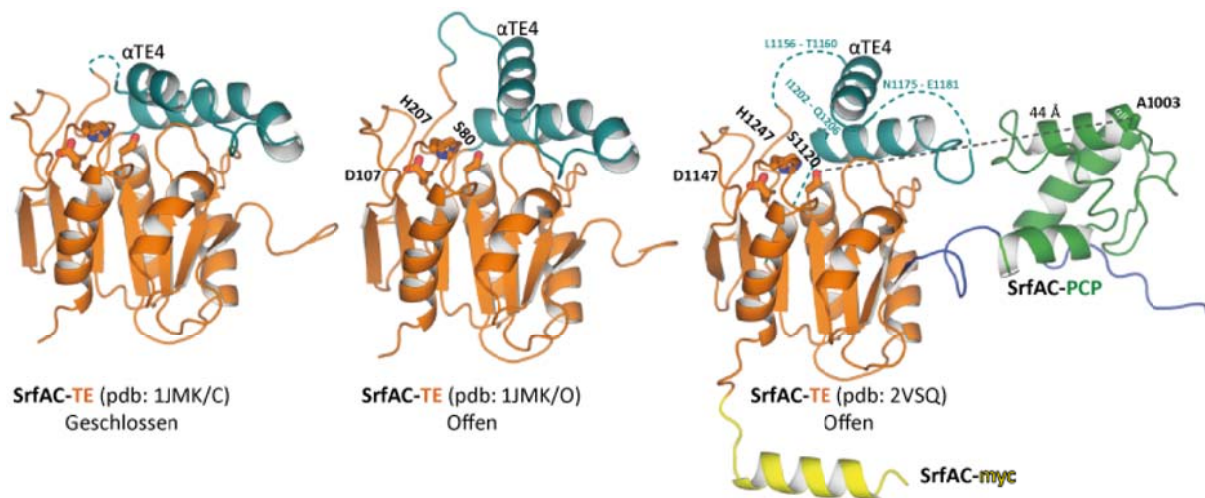


Abbildung 35: Surfactin Thioesterase zeigt keine Wechselwirkungen mit der PCP-Domäne. Im Vergleich mit der alleinstehenden SrfTE-Domäne in der offenen (pdb: 1JMK-O) und geschlossenen (pdb: 1JMK-C) Konformation, zeigt die Thioesterase aus SrfAC (pdb: 2VSQ) die offene Konformation. In dem SrfAC-Strukturmodell ist die Elektronendichte der TE-Domäne zwischen dem PCP-TE-Linker und der *myc*-Helix lokalisiert, es können keine weiteren Interaktionen festgestellt werden. Das Serin der katalytischen Triade (Ser<sup>1120</sup>, Asp<sup>1147</sup> und His<sup>1247</sup>) befindet sich 44 Å von dem Alanin<sup>1003</sup>, auf der  $\alpha$ II-Helix der PCP-Domäne, entfernt.

### 5.2.2 Intramodulare Wechselwirkungen

Das in vier Domänen organisierte SrfAC zeigt sowohl sequenzbedingte Wechselwirkungen mittels der Linker-Regionen als auch über die durch die Domänen ausgebildeten Oberflächen. Zwischen den vier struktur- und sequenzdefinierten Domänen befinden sich drei Linkerbereiche (Abb. 36). Der Linker zwischen der C- und A<sub>core</sub>-Domäne kann in N- und C-terminalen Bereich unterteilt werden, die durch ihre orthogonale Orientierung zueinander seine L-förmige Konformation ausbilden. Der N-terminale *Loop*-Bereich (Pro<sup>431</sup> - Arg<sup>445</sup>) stabilisiert die Interaktion zwischen diesen beiden angrenzenden Domänen, während der helikale C-terminale Bereich (Glu<sup>446</sup> - Pro<sup>465</sup>) starke Interaktionen mit der A<sub>core</sub>-Domäne aufweist (Abb. 36A). Die breite Interaktionsfläche zwischen der C- und der A<sub>core</sub>-Domäne, sowie die enge Interaktion mit dem C-A-Linker definieren den starren Bereich der *Workbench*. Im Gegensatz zu dem C-A-Linker lassen sich für den 15 Aminosäuren langen Linker (Pro<sup>955</sup> - Arg<sup>970</sup>) zwischen der A<sub>sub</sub>- und der PCP-Domäne (Abb. 36B) keine Kontakte mit der katalytischen Plattform feststellen, der Linkerbereich zeigt jedoch eine definierte Elektronendichte. Der nur 9 Aminosäuren lange Linker zwischen PCP- und TE-Domäne (Gly<sup>1041</sup> - Val<sup>1050</sup>) trennt diese beiden Domänen, die sonst keine weiteren Kontakte aufweisen (Abb. 36C). Im Vergleich zeigen die Linker-Regionen in dem SrfAC-Strukturmodell



grundsätzlich verschiedene Merkmale.

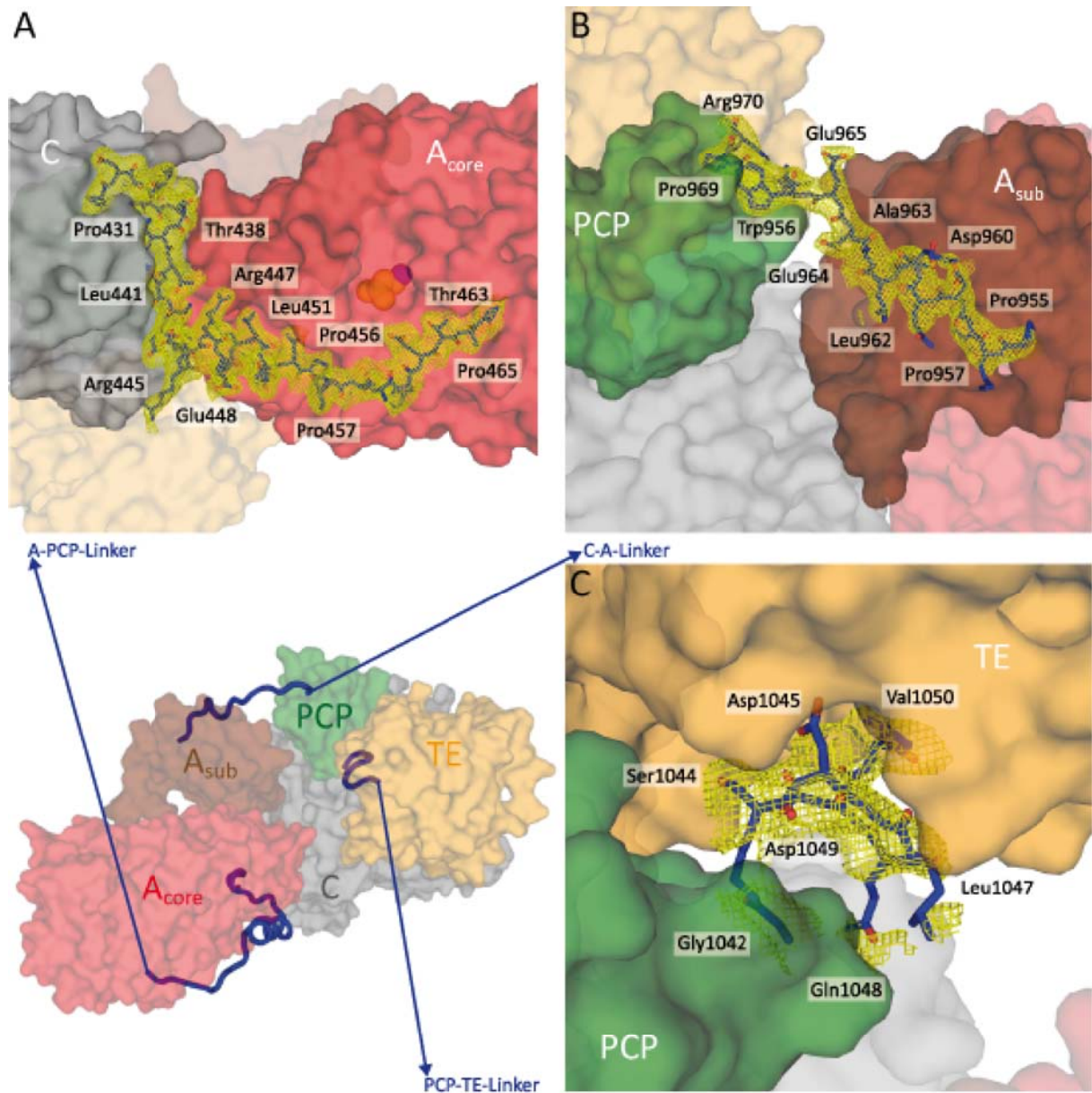


Abbildung 36: Linkerbereiche in der SrfAC-Struktur sowie deren definierte Elektronendichte. (A) Der 34 Aminosäuren lange Linkerbereich (Pro<sup>431</sup>-Pro<sup>465</sup>) zwischen der Kondensations- und der Adenylierungsdomäne (core). (B) Der 19 Aminosäuren lange Linkerbereich (Lys<sup>951</sup>-Arg<sup>970</sup>) zwischen der Adenylierungs- (sub) und der PCP-Domäne. (C) Der 9 Aminosäuren lange Linkerbereich (Gly<sup>1041</sup>-Val<sup>1050</sup>) zwischen der PCP- und der Thioesterase-Domäne. Die Elektronendichte wird bei einem  $\sigma$ -Level von 2 und Wirkungsradius von 2 Å angezeigt.

Dabei zeigt sich der signifikanteste Kontrast beim Vergleich des C-A-Linkers und des A-PCP-Linkers. Der C-A-Linker unterstützt durch seine Form sowie die vielen Interaktionen die starre Natur der *Workbench*, bestehend aus der C- und A-Domäne. Der flexible A-PCP-Linker zeigt dagegen keine signifikanten Wechselwirkungen mit der Workbench. Dabei impliziert die

Form des A-PCP-Linkers, dass die A<sub>sub</sub>-Domänenbewegung während des katalytischen Zyklus der A-Domäne die sterische Kontrolle über die Bewegung der PCP-Domäne entlang der Akzeptor-Seite der C-Domäne übernimmt. Sofern die PCP-Domäne aufgrund fehlender Beladung in ihrer Affinität zu der Akzeptor-Seite der C- oder der TE-Domäne nachlässt, wird eine Reorganisation der A<sub>sub</sub>- und PCP-Domäne und die damit einhergehende Adenylierung einer neuen Aminosäure eingeleitet. Neben der Interaktion über die Linker-Regionen wechselwirken die Domänen auch mittels ihrer Oberflächen.

Die einzelnen Aminosäuren auf der Oberfläche der jeweiligen Domäne sowie ihre Interaktionspartner sind in folgendem Diagramm, erstellt durch den CMA-Webserver[226] (*contact map analysis*), dargestellt, dabei ist die Interaktionsfläche entsprechend ihrer Ausdehnung farblich markiert (Abb. 37). Für die Interaktion wichtige Aminosäuren können so anhand der großen Interaktionsfläche (Farbe) und vertikalen bzw. horizontalen Ausdehnung entlang der Reihen bzw. Zeilen identifiziert werden. Das C-A-Interface zeichnet sich durch eine Vielzahl an Interaktionen aus, die eine 1621 Å<sup>2</sup> umfassende Fläche vom Lösungsmittelzugang abschirmen (Abb. 37A). Die zentrale Rolle für die Interaktion zwischen der C- und der A<sub>core</sub>-Domäne, sowie mit dem C-A-Linker übernimmt der zwischen den βA16 und βA17 befindliche Loop (Lys<sup>791</sup> bis Gly<sup>804</sup>). Dieser befindet sich unmittelbar vor dem Core-Motiv A6 (Gly<sup>804</sup> bis Val<sup>818</sup>) und interagiert mit der C-terminalen α-Helix αC13 (Arg<sup>416</sup> und Gln<sup>420</sup>) sowie mit dem zwischen αC8 und βC9 liegenden Loop (Leu<sup>265</sup> bis Ser<sup>272</sup>). Einzelne Aminosäuren zwischen dem Core-Motiv A7 (Tyr<sup>840</sup> und Leu<sup>845</sup>) und A8 (Gly<sup>858</sup> und Lys<sup>878</sup>) flankieren das C-A-Interface (Abb. 38A). Durch diese starken und zahlreichen Interaktionen wird eine Interaktionsfläche ausgebildet, die während des Katalysezyklus bestehen bleibt und vermuten lässt, dass die Workbench in Lösung aufrecht erhalten werden kann.

Eine weitere Reihe an Wechselwirkungen entsteht zwischen der C- und A<sub>sub</sub>-Domäne (Abb. 37A). Dabei interagieren die Reste aus dem β-Loop (Abb. ) des Core-Motiv A8 (Lys<sup>865</sup> bis Arg<sup>870</sup>; βA22-βA23) sowie aus dem Loop βA24-βA25 (Arg<sup>895</sup> bis Ala<sup>901</sup>) mit dem Loop αC6-βC8 des C-terminalen Bereichs der C-Domäne (Glu<sup>207</sup> bis Thr<sup>213</sup>; Abb. 38B). Da der β-Loop während des Katalysezyklus eine gravierende Orientierungsänderung beschreibt, können diese Interaktionen nicht durchgehend aufrecht erhalten werden. Obwohl die A<sub>sub</sub>-Domäne während des Katalysezyklus diese signifikante Orientierungsänderung vollzieht, bleibt sie

aufgrund des nur vier Aminosäuren langen Scharniers (Core-Motiv A8) eng mit der Workbench ( $A_{\text{core}}$ -Domäne) verbunden. Die Holo-PCP-Domäne dagegen muss in ihrem Katalysezyklus zwischen drei relevanten, ortsunabhängigen Orientierungen variieren.

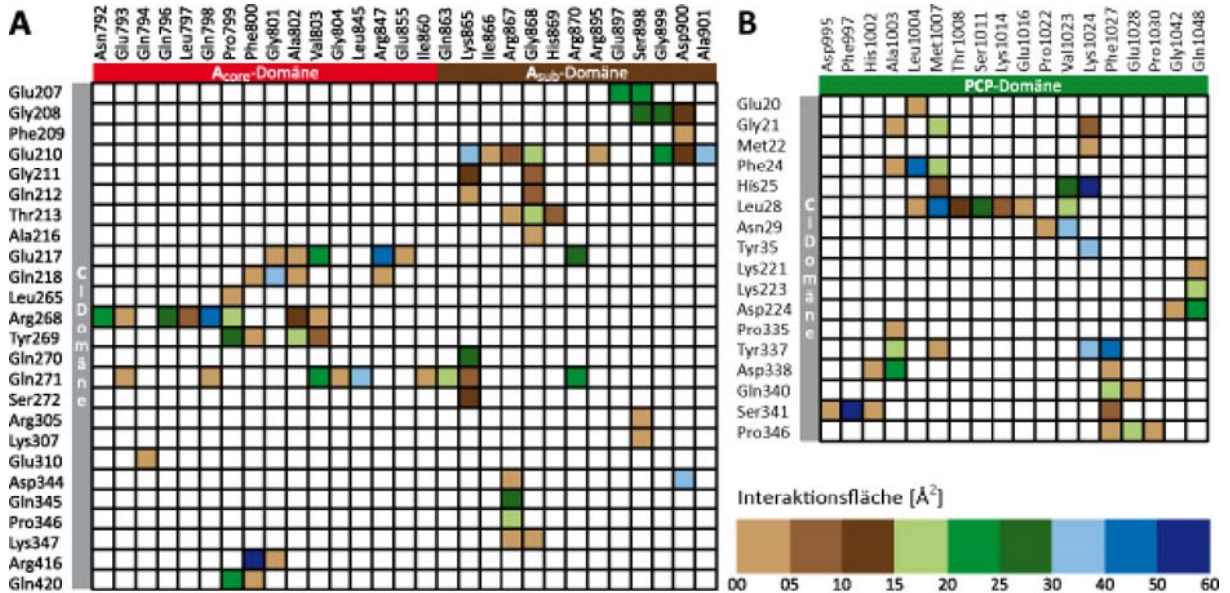
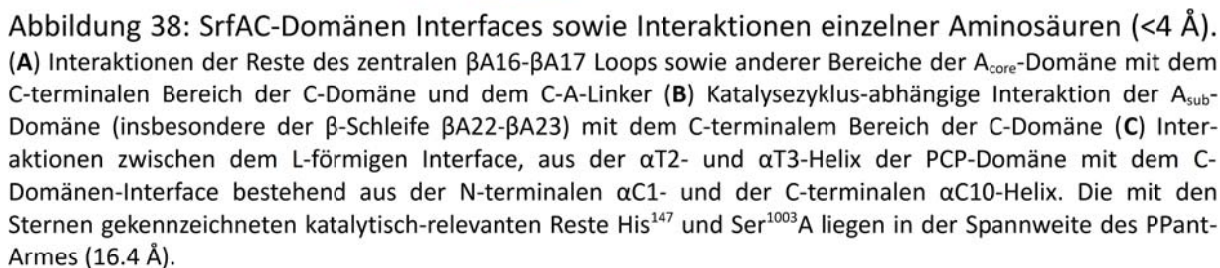


Abbildung 37: Nach Aminosäuren aufgelöste Interaktionsflächen der SrfAC-Domänen. (A) 1621 Å<sup>2</sup> umfassende Workbench-Interaktionsfläche zwischen der C-Domäne und der  $A_{\text{core}}$ - sowie  $A_{\text{sub}}$ -Domäne. (B) 975 Å<sup>2</sup> umfassende Interaktionsfläche zwischen der C- und PCP-Domäne.

Im Zusammenspiel mit der Konformationsänderung der  $A_{\text{sub}}$ -Domäne ermöglicht die erste der drei Orientierungen die Aminoacylübertragung auf den PPant-Arm der PCP-Domäne. Das im Strukturmodell vom variierten Ala<sup>1003</sup> 57.3 Å entfernte Leucin würde dann als Thioester auf den PPant-Arm übertragen werden. In der zweiten Orientierung wird das Leucin an der C-Domäne um die vorangehenden Elongationsprodukte erweitert. Das für diesen Schritt katalytisch verantwortliche His<sup>147</sup>, im N-terminalen Bereich der C-Domäne lokalisiert, ist Bestandteil des Core-Motivs C3 und befindet sich in der Reichweite des PPant-Arms (16.4 Å; Abb. 38C).

In der dritten Orientierung wird das lineare Elongationsprodukt, das als Thioester-Intermediat am PPant-Arm gebunden ist, auf das durch die katalytische Triade (Ser<sup>1120</sup> - Asp<sup>1147</sup> - His<sup>1247</sup>) deprotonierte, 43.5 Å von Alanin<sup>1003</sup> entfernte Serin<sup>1120</sup> als Oxoester-Intermediat übertragen. Anschließend würde durch eine interne Zyklisierungsreaktion die Freisetzung des Produktes Surfactin erfolgen. Bedingt durch die Serin<sup>1003</sup> zu Alanin Variation, kann der PPant-Arm nicht an die PCP-Domäne übertragen werden, dennoch orientiert sich die





Helix und schirmen so eine 975 Å<sup>2</sup> umfassende Fläche vom Lösungsmittelzugang ab (Abb. 37B). Wie bereits bei EntB-ArCP (*aryl carrier protein*) für die Interaktion mit EntF aus der Enterobactin Biosynthese gezeigt wurde[110], bilden insbesondere Methionin<sup>1007</sup> (αT2) und Phenylalanin<sup>1027</sup> (αT3) hydrophobe Interaktionen mit Phenylalanin<sup>24</sup> und Leucin<sup>28</sup> aus der αC1- und Tyrosin<sup>337</sup> aus der αC10-Helix (Abb. 38C). Neben der hydrophoben Wechselwirkung zeigt Lysin<sup>1027</sup> (Ser<sup>1003</sup>Ala + 21aa) starke polare Interaktionen mit Histidin<sup>25</sup> (αC1) und Tyrosin<sup>35</sup>. Das Lyssin<sup>66</sup> (Ser<sup>45</sup> + 21aa) aus der Holo-TycC<sub>3</sub>-PCP zeigt bei dem Pendeln zwischen der AH- und H-Konformationen eine der höchsten Differenzen im Signal des TROSY-HSQC Spektrums[227] und deutet bereits auf die Beteiligung des Lysin<sup>66</sup> (αT3) an PCP-Interaktionen mit anderen Domänen hin.

### 5.2.3 Intermodulare Wechselwirkung

Unerwartet zeigte das Strukturmodell eine wohldefinierte Interaktion zwischen dem durch β-Stränge geformten, N-terminalen Teil der C-Domäne und einem helikalen Strukturelement des benachbarten Moduls, das sich aus dem C-terminalen *myc*-Epitop[142], sowie dem His<sub>6</sub>-Tag-Bereich zusammensetzt (Ser<sup>1275</sup> bis Ala<sup>1290</sup>). Durch diese Interaktion bilden die Moleküle entlang der X,Z-Ebene des Kristallgitters (Abb. 39A) einzelne Glieder einer langen Kette. Durch Sequenzvergleiche konnte bereits gezeigt werden, dass die nichtribosomalen Synthetasen selektiv miteinander kommunizieren, um die Synthesereihenfolge aufrecht erhalten zu können[133]. Das Interaktionsprinzip beruht dabei auf komplementären Partnern, die durch spezifische Wechselwirkungen die synthetasenübergreifende Synthesereihenfolge garantieren. Der helikale C-Terminus der E-Domäne, die COM-Helix, interagiert dabei mit der folgenden C-Domäne, anschließend erfolgt die Elongationsreaktion. Im Gegensatz zu der postulierten Helix-Helix-Interaktion[133, 228] zeigt das SrfAC-Strukturmodell ein komplexeres Motiv, die COM-Hand, die als Andockstelle für die COM-Helix dient. Das COM-Hand Motiv wird durch drei β-Faltblätter (βC1-βC3-βC4) geformt, die multiple Kontakte mit den Resten der artifiziellen helikalen COM-Helix ausbilden (Abb. 39B). Dabei wechselwirken die hydrophoben Reste der zentralen Bodenfläche der COM-Hand (Phe<sup>4</sup>, Tyr<sup>14</sup>, Ile<sup>71</sup>, Val<sup>75</sup>, Val<sup>79</sup> und Val<sup>81</sup>) mit denen der zugewandten Seite der COM-Helix (Phe<sup>1279</sup>, Val<sup>1283</sup> und Leu<sup>1287</sup>).



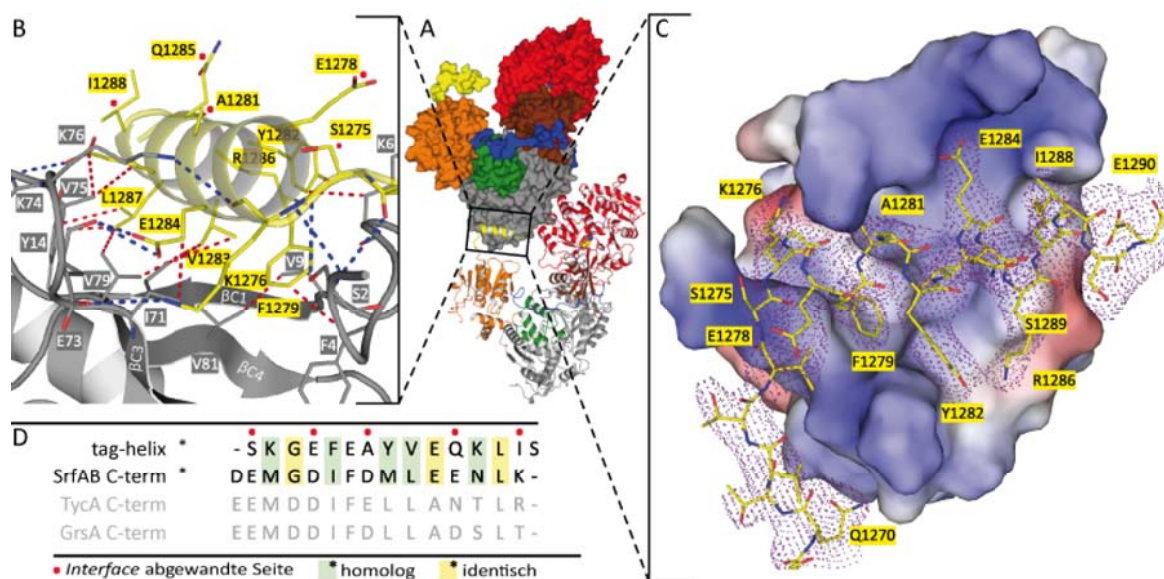


Abbildung 39: Durch die artifizielle COM-Helix besetztes COM-Hand-Interface aus SrfAC (A) Kristallpackung entlang der X,Z-Ebene, angedeutet durch die Interaktion zweier SrfAC-Moleküle. (B) Multiple hydrophobe (rot) und polare (blau) Interaktionen zwischen den einzelnen Aminosäuren der COM-Helix (gelb) und der COM-Hand (grau). (C) Oberflächenrepräsentation der elektrostatischen Wechselwirkungen zwischen dem Protein (COM-Hand) und ionischem Lösungsmittel (Wasser) nach der Poisson-Boltzmann-Gleichung (zwischen  $-5$  (rot) und  $+5$  (blau)  $k_B T/e_c$ ), sowie die Elektronendichterepräsentation (isodot) und Strukturmodell der artifiziellen COM-Helix (gelb). Die COM-Hand zeigt in der „Handfläche“ ein hydrophobes Interface, während die Seiten polare Interaktionen bevorzugen. (D) Sequenzalignment der kompatiblen COM-Helices. Alle Sequenzen zeigen eine hohe Aminosäure-Homologien für die dem Interface zugewandte Seite.

Zusätzlich werden die Interaktionen durch die seitlichen polaren Wechselwirkungen zwischen COM-Helix (Ser<sup>1275</sup>, Lys<sup>1276</sup> und Glu<sup>1284</sup>) und COM-Hand (Ser<sup>2</sup>, Glu<sup>73</sup>, Lys<sup>74</sup> und Lys<sup>76</sup>) erweitert. Die elektrostatischen Oberflächenrepräsentationen der COM-Hand wurden nach der Poisson-Boltzmann-Gleichung berechnet und anschließend durch die Software APBS-Tools2[229] und PyMol 1.3[230] visualisiert. Dabei zeigt die Oberflächenrepräsentationen ein positives Potential (blau) für den peripheren Bereich, sowie ein hydrophobes (weiß) Interface für die zentrale, der COM-Helix zugewandte, Region (Abb. 39C). Die Affinität der COM-Helix zu der korrespondierenden COM-Hand ermöglicht nicht nur die konsequente Einhaltung der Synthesereihenfolge, sondern auch eine artifizielle Rekombination der Synthetasen[155]. Die funktionelle sowie sequenzielle Homologie zwischen den Initiationsmodulen der Gramacidin S und Tyrocidin A produzierenden NRPS-Systeme ermöglichten bereits die Kreuzkupplung der korrespondierenden Synthetasen[231]. Ein Sequenzvergleich des COM-Helix-Motivs mit SrfAB zeigt eine hohe Homologie für die der COM-Hand zugewandten Seite (Abb. 39D)[155]. Im Experiment konnte die Bildung des D-Phe-L-Leu-Dipeptids durch die Kreuzkupplung

(*cross-talk*) zwischen TycA (1088 Aminosäuren, 123 kDa) und SrfAC (1274 Aminosäuren, 144 kDa) nachgewiesen werden (Abb. 40).

Nach Inhibition der TE-Domäne durch den Protease-Inhibitor Phenylmethylsulfonylfluorid (PMSF) konnte keine Produktbildung beobachtet werden, das Dipeptid wurde nicht durch hydrolytische sondern durch die katalytische Abspaltung durch die TE-Domäne freigesetzt. Neben dem <sub>DL</sub>-Hauptprodukt (<sub>D</sub>-Phe-<sub>L</sub>-Leu) konnte ebenfalls das entsprechende <sub>LL</sub>-Diastereomer (<sub>L</sub>-Phe-<sub>L</sub>-Leu) durch Vergleichsmessungen mit Standards in geringen Mengen identifiziert werden[155]. Bereits eine Deletion der drei C-terminalen Aminosäuren aus dem COM-Helix-Bereich von TycA führt zum Verlust der Fähigkeit der intermodularen Wechselwirkung. Dies deutet auf die essentielle Rolle des COM-Helix-COM-Hand-Motivs für die Aminosäure- bzw. Peptidylübertragung zwischen den interagierenden Synthetasen hin[133]. Um die COM-Helix-COM-Hand-Interaktion zu untersuchen, wurden den Interaktionsassays die in der Sequenz mit der korrespondierendem COM-Helix-Motiv identischen Peptide (COM-Peptide) zugegeben. Durch die Zugabe der COM-Peptide (Tab. 15) sollte der Interaktionspartner der COM-Helix, die COM-Hand, besetzt werden, um so allosterisch eine Interaktion zwischen TycA und SrfAC zu inhibieren. Erste Messungen mit dem *myc*-Peptid zeigten nur eine geringe Inhibition.

Name	myc (15AA)	Inverse myc	SrfAB (15AA)	TycA (23AA)
Sequenz	SKGEF <b>EAYVE</b> QKLIS	SILKQEVYAEFEGKS	DEMGD <b>IFD</b> MLEENLK	FSVKGLQ <b>MEEMDD</b> IFELLANTLR
Sekundärstrukturwahrscheinlichkeit (> 66%): AA - αHelix   AA - β-Strang   AA - Coil				

Tabelle 15: Sekundärstruktur-Vorhersage für die Sequenzen der COM-Peptide.

Aufgrund der Unlöslichkeit der COM-Peptide in wässrigen Lösungen und dem Aktivitätsverlust bei höherer Dimethylsulfoxid-Zugabe (DMSO) ist eine quantitative Aussage nicht möglich, es lassen sich jedoch Tendenzen feststellen. Die aktive helikale Form ist unter anderem von der Länge des COM-Peptides abhängig, da nach Berechnungen (*umbrella sampling*) besonders der C-terminale Bereich bei Raumtemperatur nicht die α-helikale Form bevorzugt[232-234].

Dabei können, abhängig von der Ladung des Peptides, sowie der sequenzbedingten Geometrie, bis zu 42% der Aminosäuren in ein helikales Sekundärstrukturelement eingebunden werden[232]. Eine Sekundärstruktur-Vorhersage für die COM-Peptide (Tab. 15),

wie auch für das Peptid mit der inversen *myc*-Sequenz, ergab für das TycA-COM-Peptid die beste Vorhersage[235]. Experimentelle Untersuchungen mit dem TycA-COM-Peptid zeigen einen Einfluß auf die Bildung des Dipeptids durch die Kreuzkupplung zwischen TycA und SrfAC. Dabei sinkt die Dipeptid-Produktion auf 56% bei einer 0.5%-igen DMSO-Konzentration und einem 100-fachen Überschuss (Abb. 41).

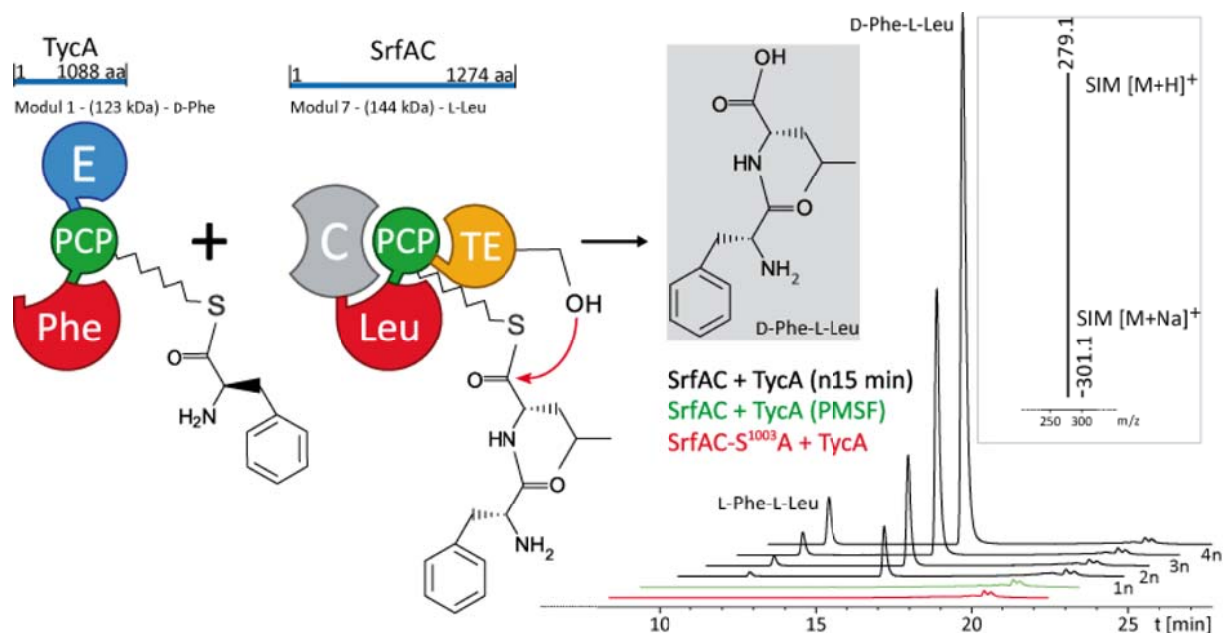


Abbildung 40: Kreuzkupplung (cross-talk) zwischen TycA [AT<sub>Phe</sub>E] und SrfAC [CA<sub>Leu</sub>TTE]. Das rot eingefärbte Chromatogramm (SrfAC-S1003A + TycA) zeigt keine D-Phe-L-Leu Bildung, da dem SrfAC-S1003A die Fähigkeit zur kovalenten Bindung des PPant-Arms fehlt. Durch die Inhibition der Ser<sup>1120</sup> in der TE-Domäne kann ebenfalls keine Dipeptidbildung beobachtet werden (grün). Der zeitliche Verlauf (15, 30, 45 und 60 min) zeigt die zunehmende Bildung von D-Phe-L-Leu, sowie in kleinen Mengen das nicht epimerisierte L-Phe-L-Leu (schwarz). SIM (single ion mass) der beiden Diastereomere [M+H]<sup>+</sup>, sowie der Natrium-Komplexe [M+Na]<sup>+</sup> wurden zeitlich aufgelöst (Chromatogramm).

Auch wenn die Erhöhung der TycA-COM-Peptid-Konzentration (1:200) die Produktion weiter auf 44% sinken lässt, zeigen die Messungen bei 5%-iger DMSO-Konzentration, dass der Einfluß des Peptids durchaus signifikant ist. Die Zugabe der anderen COM-Peptide, sowohl des nativen Interaktionspartners aus SrfAB als auch eines nicht zu helikalen Strukturen neigendem Kontroll-Peptides mit der inversen *myc*-Sequenz, zeigen bei diesen Konzentrationen keinen Einfluß auf die Dipeptid-Produktion. Die Konzentrationserhöhung auf 5% DMSO zeigt zwar deutlich erhöhte Löslichkeit der COM-Peptide in der wässrigen Lösung und somit eine stärkere Inhibition in Abhängigkeit von der Peptidkonzentration, die Enzymaktivität sinkt jedoch auf unter 10% der nativen Produktionsrate. Der Einfluss des TycA-COM-Peptids auf die Dipeptid-Produktion ist für beide Konzentrationen messbar, da

jedoch die Konzentration der  $\alpha$ -helikalen aktiven Form des Peptides in der Lösung nicht zu bestimmen ist, kann eine quantitative Auswertung nicht stattfinden. Jedoch kann festgestellt werden, dass das TycA-COM-Peptid die Dipeptid-Produktion empfindlich stört und durchaus einen Einfluß auf die Interaktion zwischen TycA und SrfAC nimmt.

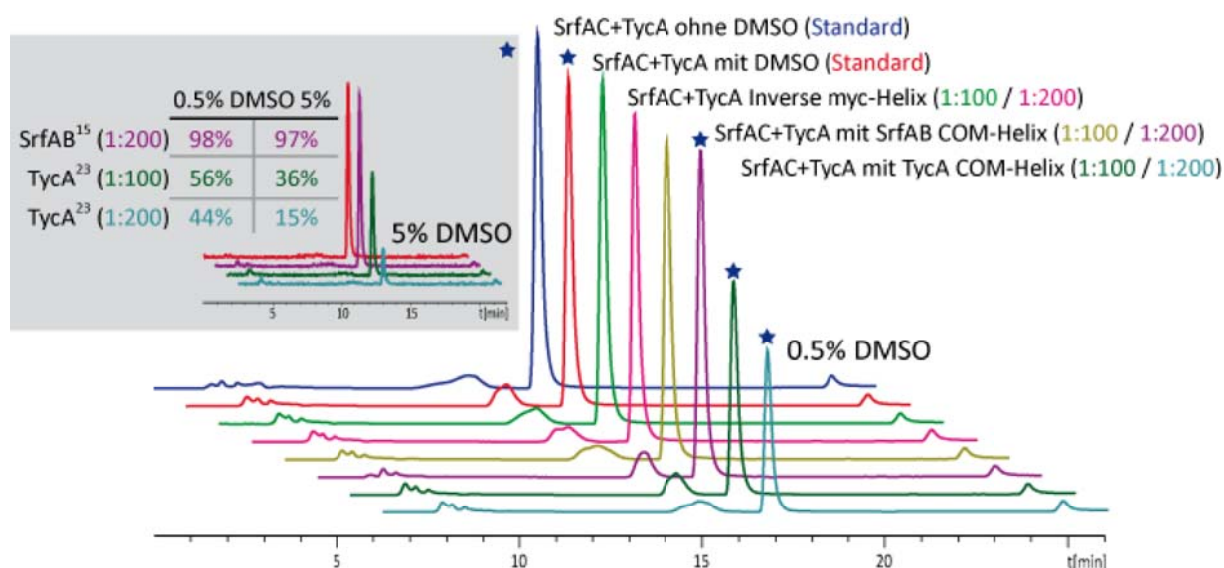


Abbildung 41: Allosterische Inhibition durch die mit der entsprechenden COM-Helix homologen Peptide (*inverse myc* (Kontrolle), SrfAB und TycA) bei 0.5% und 5%\* DMSO.

## 5.3 Modell multimodularer NRPS-Systeme

Bisher bekannte nichtribosomale Peptidsynthetasen können aus bis zu zwölf Modulen sowie einer Masse von bis zu 1.6 MDa bestehen und gehören so zu den größten Enzymen lebender Systeme[86]. Frühere Modelle beschreiben die multimodularen Synthetasen durch ein vereinfachtes „Perlenschnur“-Modell (*pearl-on-a-string model*), bei dem die einzelnen Domänen durch flexible Linker verbunden sind und so freie räumliche Orientierungen zueinander einnehmen können. Somit würde die sterische Kontrolle über die Reaktionsabfolge sowie Interdomänen-Wechselwirkungen dem Beladungszustand der PCP-Domäne unterliegen. Das SrfAC-Strukturmodell ermöglicht durch die definierten Orientierungen einzelner Domänen, Rückschlüsse auf die Funktionsweise der modularen NRPS-Systeme zu ziehen. Dabei zeigt besonders die starke Interaktion zwischen der C- und A<sub>core</sub>-Domäne und die damit einhergehende Ausbildung der *Workbench* die Komplexität der Funktionsweise in NRPS-Systemen. Besonders die Kombination zwischen der starren *Workbench* und den

dynamischen A<sub>sub</sub>- und PCP-Domänen sowie deren Reorganisation während der Reaktionszyklen zeigen die Vielfältigkeit im Wirkungsablauf der NRPS-Systeme. Gleichzeitig ermöglicht die Kombination der TycC<sub>5-6</sub>-PCP-C-[121] und SrfAC[236]-Strukturmodelle den Zugang zu einem strukturbasierten multimodularen NRPS-Modell (Abb. 42A). Durch die strukturelle Überlagerung (Superposition) der C-Domäne aus beiden Strukturmodellen entsteht ein durch PCP-Domänen flankiertes Terminationsmodul (Abb. 42B). Die Deletion der TE-Domänen führt zu einem Elongationsmodell, das durch repetitive Superposition mit dem Terminationsmodul die Konstruktion des multimodularen NRPS-Modells ermöglicht (Abb. 42C). Das so erzeugte Modell beschreibt eine einzigartige schraubenförmige globale Struktur. Dabei liegen die C- und PCP-Domänen entlang der Achse und bilden die Reaktionsachse (*assembly line*), während die A-Domäne den peripheren Bereich bildet. Die einzelnen Module werden im 120°-Winkel gegen den Uhrzeigersinn im Abstand von 45 Å voneinander translatiert. Die maximale Auslenkung zwischen der zentralen PCP-Domäne (Ser<sup>1003</sup>Ala) und der peripheren A-Domäne (Glu<sup>594</sup>) beträgt 98 Å, die Ganghöhe, bei der die Module eine volle Umdrehung beschreiben 127 Å. Durch die Superposition der PCP-Domäne aus beiden Strukturmodellen zeigt das Modell den *acceptor site only* Zustand an. Bei diesem Akzeptor-Modell interagiert die PCP-Domäne über den αTI-αTII-Loop und die αTII-Helix mit der *downstream* C-Domäne (*acceptor site*). Mit der *upstream* C-Domäne zeigt die PCP-Domäne eine unproduktive doch signifikante Interaktion mit dem N-terminalen Ende der αTI-Helix[121]. Dabei ist die Interaktionsfläche deckungsgleich zu der ACP-Domäne aus der 6-Deoxyerythronolid B PKS-Synthase. Diese zeigt jeweils zwei Interaktionsflächen, für die Elongation (αTI-αTII-Loop und αTII-Helix) und dem Transfer (N-terminales Ende von der αTI-Helix)[72]. Die übergeordnete NRPS-Struktur mit einer dreizähligen Schraubachse basiert ausschließlich auf verfügbaren Strukturdaten und zeigt mannigfaltige Vorteile. Die abwechselnde Reihenfolge der PCP- und C-Domänen entlang der Reaktionsachse (*assembly line*) ermöglicht eine Abschirmung der Zwischenprodukte vor der proteolytischen Aktivität aus dem Cytosol. Zusätzlich sind die Reaktionskammern der einzelnen C-Domänen für die PCP-gebundenen PPant-Arme zugänglich und gewähren zugleich den Schutz der Intermediate vor Hydrolyse. Gleichzeitig ermöglichen die nach außen gerichteten *Workbench*-Bereiche durch exponierte, periphere Orientierung der A-Domänen den Zugang sowohl der



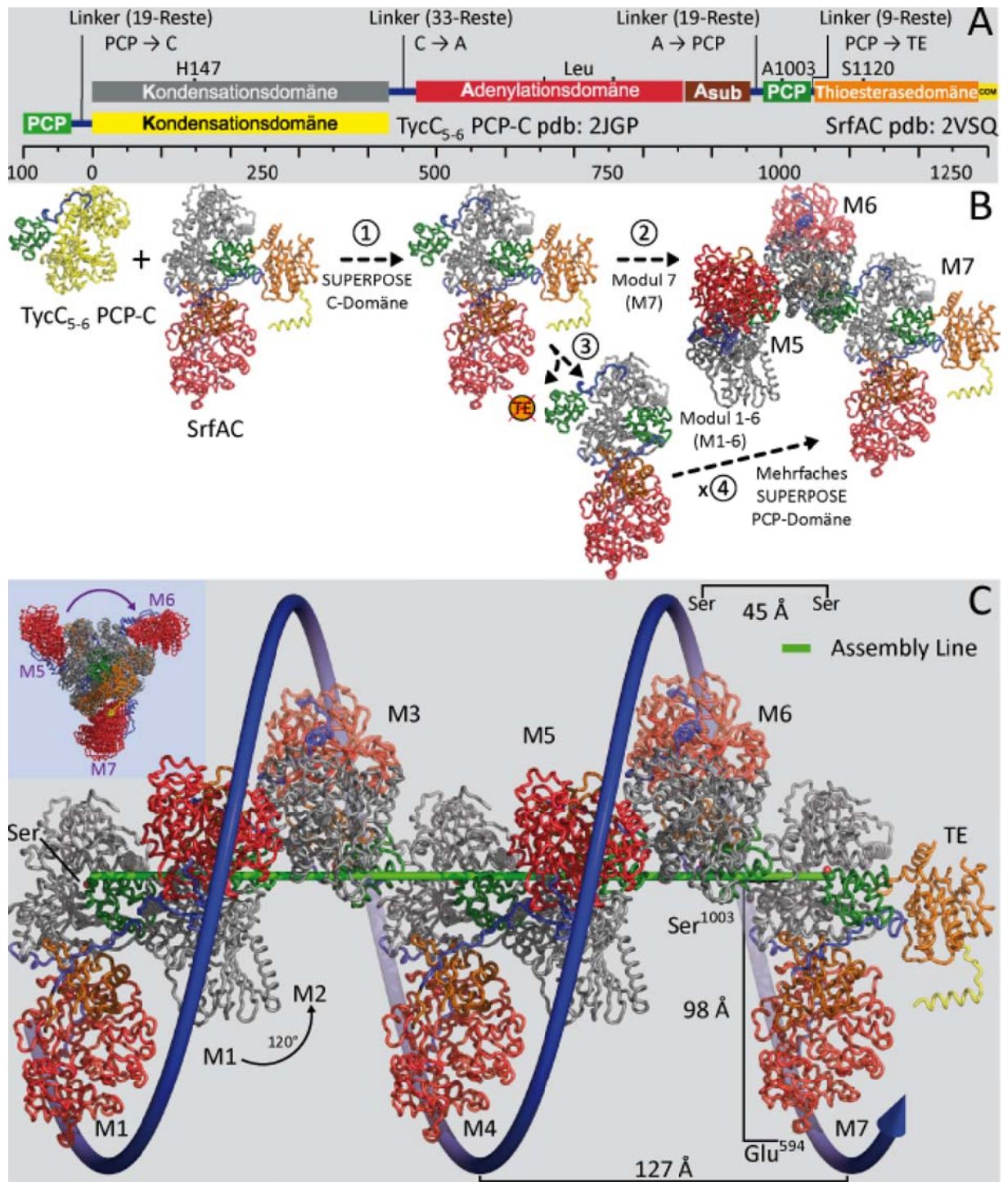


Abbildung 42: Konstruktion und Aufbau multimodularer NRPS als „acceptor site“-Modell. (A) Schematische Darstellung von TycC<sub>5-6</sub> PCP-C und SrfAC C-A-PCP-TE, den Linkerbereichen sowie der übereinstimmenden Region für das Akzeptor-Modell. (B) Konstruktionsroute für das multimodulare NRPS-Strukturmodell basierend auf den Strukturmodellen von TycC<sub>5-6</sub> (pdb: 2JGP) und SrfAC (pdb: 2VSQ; ①). Das Modell wird durch repetitive Superposition (④) des, durch die Deletion der TE-Domäne (③) resultierenden, Elongationsmoduls mit dem Terminationsmodul bzw. Modell (②) konstruiert. (C) Der Aufbau einer repräsentativen heptamodularen Synthetase. Die Anordnung der einzelnen Module beschreibt eine Schraube mit einer Drehung von 120° pro Modul gegen den Uhrzeigersinn. Die Ganghöhe der so beschriebenen Schraube beträgt 127 Å pro Drehung, dabei wird das Modul entlang der Achse (*assembly line*) um 45 Å translatiert, während die maximale Auslenkung (Glu<sup>594</sup>) 98 Å beträgt.

*in trans* wirkenden Tailoring-Enzyme, als auch die Inkorporation variabler *in cis* agierender Modifikationsdomänen zwischen den Core-Motiven A8 und A9. Die Natur des Modells würde ohne globale Strukturstörungen eine lokale Reorientierung der PCP-Domäne erlauben, um so einen von dem „*acceptor site only*“ - Modell abweichenden Zustand beschreiben zu können.

### 5.4 Probenpräparation und Analyse

Die Arretierung der PCP-Domäne schien der Schlüssel für die Strukturlösung des SrfAC-S1003A-Strukturmodells zu sein. Dabei interagiert die arretierte PCP-Domäne von der Akzeptor-Seite mit der C-Domäne des Moduls. Um ein globales multimodulares NRPS-Strukturmodell erstellen sowie die einzelnen Stationen der NRPS-Synthese untersuchen zu können, wurde der Beladungszustand der PCP-Domäne variiert (Tab. 16). Die Beladungsvarianten der PCP-Domäne sind durch die Synthesereihenfolge der nichtribosomalen Peptidsynthetase bestimmt und somit durch die entsprechende A-Domäne gesteuert. Den Beladungszustand der zu untersuchenden PCP-Domäne über die Aktivität der entsprechenden A-Domäne zu verfolgen, ist wenig zuverlässig. Durch den ATP/PP<sub>i</sub>-Austausch als Aktivitätsindikator der entsprechenden A-Domäne kann nur die Aktivität einzelner A-Domänen überprüft werden. Jedes Modul in multimodularen NRPS-Systemen enthält in der Regel eine A-Domäne, die für die Aktivierung und Thiolierung der korrespondierenden Aminosäure verantwortlich ist. Für den Fall, dass zwei oder mehrere A-Domänen einer nichtribosomalen Peptidsynthetase die gleiche Aminosäure aktivieren, wie z.B. bei TycB<sub>2-3</sub> (l-Phe) oder SrfAA<sub>2-3</sub> (l-Leu), kann eine zuverlässige Aussage über die Aktivität der einzelnen A-Domäne durch den ATP/PP<sub>i</sub>-Austausch nicht getroffen werden, der Beladungszustand der PCP-Domänen kann dabei nicht untersucht werden. Da die korrespondierende Aminosäure als Thioester am PPant-Arm der PCP-Domäne gebunden ist, kann der Beladungszustand mit nativem PPant-Arm unter den basischen Bedingungen für den tryptischen Verdau wegen der hohen Hydrolyse-Anfälligkeit nicht verfolgt werden. Aus diesem Grund wurde für analytische Zwecke der gekoppelte Beladungsassay als Aktivitätsindikator angewendet[173]. Für eine 100%-ige Beladungspopulation im präparativen Maßstab wurde dieser Assay optimiert.

Die weiteren makromolekularen Untersuchungen können jedoch nur dann zuverlässig ausgewertet werden, wenn die Population der unerwünschten Beladungszustände unter der

Nachweisgrenze liegt. Für die Kristallisationsansätze sowie die Untersuchungen durch SAXS und CryoEM wurden die Beladungszustände der einzelnen PCP-Domänen zwischen der Apo- und Holo-Form mit Thiol- und Amino-Terminus sowie zusätzlich verknüpften Derivaten variiert (Tab. 16).

Protein- bezeichnung	Domänen- Organisation, Beladungs- zustand	Vektor	Ursprung	Expression	Aufreinigung	Screening	Kristallwachstum	Diffraction	Optimierung	SAXS	CryoEM
Unimodulare nichtribosomale Peptidsynthetasen											
GrsA	[ATE] <sup>a</sup> , -sh	pQE60	<i>B. brevis</i> ATCC 9999	✓	✓	✓	✗	■	■	■	■
GrsA <sup>S573A</sup>	[ATE]	pQE60	<i>B. brevis</i> ATCC 9999	✓	✓	✓	✗	■	■	■	■
TycA	[ATE] <sup>a</sup> , -sh, -nh2, -nh-aa	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	✓	✗	■	■	■	■
TycA <sup>S562A</sup>	[ATE]	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	✓	✗	■	■	■	■
TycB <sub>1</sub>	[CAT] <sup>a</sup> , -sh	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	✓	✗	■	■	■	■
TycB <sub>1</sub> <sup>S1007A</sup>	[CAT]	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	✓	✗	■	■	■	■
TycB <sub>3</sub>	[CATE] <sup>a</sup> , -nh2, -nh-aa	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	■	■	■	■	✓	■
TycB <sub>3</sub> ΔE	[CAT] <sup>a</sup>	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	■	■	■	■	✓	■
SrfAC	[CATTE] <sup>a</sup> , -sh, -nh2, -nh-aa	pTrcHis2	<i>B. subtilis</i> ATCC 21332	✓	✓	✓	✗	■	■	✓	■
SrfAC <sup>S1003A</sup>	[CATTE]	pTrcHis2	<i>B. subtilis</i> ATCC 21332	✓	✓	✓	✓	✓	✓	■	■
Dimodulare nichtribosomale Peptidsynthetasen											
BacA <sub>1-2</sub>	[ATCyAT] <sup>a</sup>	pQE70	<i>B. licheniformis</i> ATCC 10716	✓	✓	✓	✗	■	■	■	■
TycB <sub>2-3</sub>	[ATCATE] <sup>a</sup> , -nh2, -nh-aa	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	✓	✗	■	■	✓	■
TycB <sub>2-3</sub> ΔE	[ATCAT] <sup>a</sup>	pQE60	<i>B. brevis</i> ATCC 8185	✓	✓	■	■	■	■	✓	■
TycB <sub>2-3</sub> FL	[CATCATE] <sup>a</sup> , -nh2, -nh-aa	pET200/D	<i>B. brevis</i> ATCC 8185	✓	✓	✗	■	■	■	✓	✓
LgrD <sub>1-2</sub> ΔE	[FATCAT] <sup>a</sup>	pBAD	<i>B. brevis</i> ATCC 8185	✓	✓	✗	■	■	■	■	■
Trimodulare nichtribosomale Peptidsynthetasen											
SrfAA	[CATCATCATE] <sup>a</sup> , -nh2, -nh-aa, -nhG	pQE60	<i>B. subtilis</i> ATCC 21332	✓	✓	■	■	■	■	■	✓
SrfAB	[CATCATCATE] <sup>a</sup>	pQE60	<i>B. subtilis</i> ATCC 21332	✗	■	■	■	■	■	■	■
-a – apo; -sh – holo (CoA), -nh2 – holo (Amino-CoA), -nh-aa – holo-Aminosäure (Amino-CoA), -nhG – holo-Gold											
<div> <span>✓</span> Erfolgreich absolviert                             <span>✗</span> Nicht beobachtet/durchgeführt                             <span>■</span> Nicht verwendet                         </div>											

Tabelle 16: Übersicht über die in dieser Arbeit untersuchten rekombinanten Proteine.

Während die Beladung durch CoA (Thiol-Terminus) *in vivo* durch die Coexpression von Gsp, der Phosphopantothein-Transferase aus dem Gramicidin S-Cluster, katalysiert wird, erfolgt die Beladung durch Amino-CoA (Amino-Terminus) *in vitro* und wurde durch Sfp katalysiert. Die Verknüpfung der korrespondierenden Aminosäure mit dem Amino-Terminus des PPant-Derivates erfolgt in multimodularen NRPS-Systemen durch die entsprechende A-Domäne. Dadurch ist es möglich, den PPant-Arm der jeweiligen PCP-Domäne auch in heterologen NRPS-Systemen mit der korrespondierenden Aminosäure zu verknüpfen. Nach dem tryptischen Verdau wurde ein Massenspektrum der zu untersuchenden Probe aufgenommen



(Abb. 43A1).

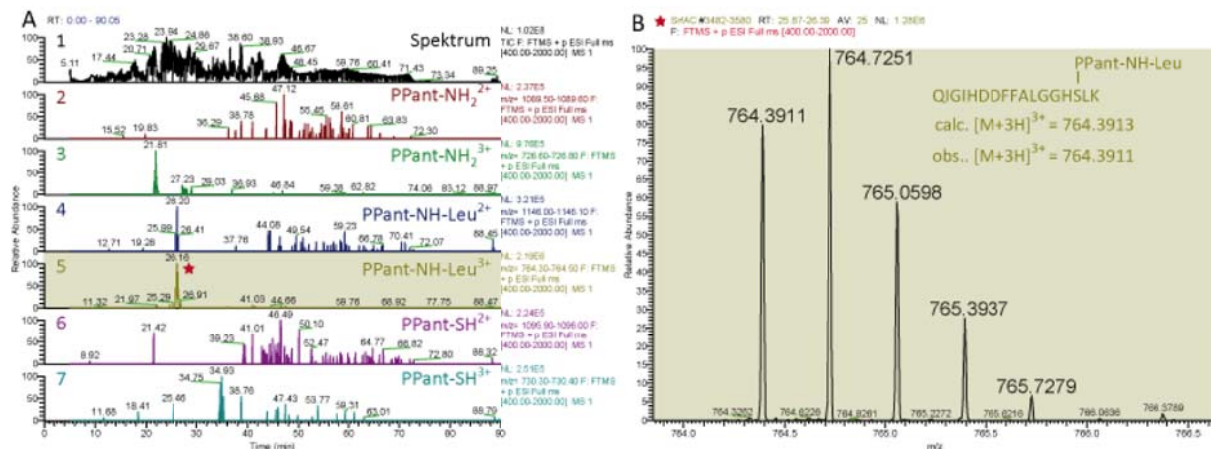


Abbildung 43: Massenanalyse des SrfAC-PCP-Fragments nach dem tryptischen Verdau. (A) Das Massenspektrum (Spektrum 1) sowie die Ionenspektren einzelner Massenbereiche. Aus Übersichts-Gründen wurden nur die  $[M+2H]^{2+}$ - und  $[M+3H]^{3+}$ -Massenspektren der einzelnen Beladungsvarianten PPant-NH<sub>2</sub> (Spektren 2 und 3), PPant-NH-Leu (Spektren 4 und 5) und PPant-SH (Spektren 6 und 7) des SrfAC-PCP-Fragments (Sequenz: QIGIHDDFFALGGHSLK) dargestellt. (B) Das Spektrum 5 des Massenbereichs  $m/z = 764.30 - 764.50$  enthielt bei einer Retentionszeit (RT) von 26.16 min die gesuchte dreifach-geladene Masse (beobachtete Masse: obs.  $[M+3H]^{3+} = 764.3911$ ; berechnete Masse: calc.  $[M+3H]^{3+} = 764.3913$ ), die geringe Abweichung ist innerhalb der gerätespezifischen Abweichung.

Über einen Filteralgorithmus wurden anschließend definierte Massenbereiche näher untersucht. Um die gesuchte Ionenmasse identifizieren zu können wurden alle Massenpeaks eines gefilterten Massenbereichs überprüft. Jede der gesuchten Massen kann als mehrfach-geladenes Ion detektiert werden, so dass bei jeder Probe alle Möglichkeiten überprüft wurden. Bei einer Beladung mit Amino-CoA und anschließender Verknüpfung mit der korrespondierenden Aminosäure wurde zuerst überprüft, ob eine *in vivo*-Beladung während der Proteinproduktion in der Zelle stattgefunden hat. Dafür wird in dem entsprechenden Massenbereich nach den geladenen Ionenmassen ( $[M+H]^+$  bis  $[M+5H]^{5+}$ ) des tryptischen Fragments mit dem entsprechenden PPant-Arm (Thiol-Terminus) gesucht. Nachdem das Protein in der Apo-Form gereinigt wurde, konnte *in vitro* der Beladungassay durchgeführt werden. Anschließend wurde durch die Suche nach den einzelnen Ionenmassen der Beladungszustand der PCP-Domäne bestimmt (Abb. 43A2-7). Nur wenn die Population der gesuchten Ionenmasse zu 100% bestätigt wurde, kann die Probe für die weitergehenden makromolekularen Strukturuntersuchungen verwendet werden (Abb. 43B).

Der Vorteil dieser Methode ist, dass nicht nur die durchschnittliche Aktivität einer A-Domäne, wie beim ATP/PP<sub>i</sub>-Austausch, sondern auch die spezifische Aktivität jeder

einzelnen A-Domäne in dem Zielprotein, sowie die Thiolierungsfähigkeit der korrespondierenden PCP-Domäne bestimmt werden kann. Somit ist es möglich die Zielproteine für die weitergehenden makromolekularen Strukturuntersuchungen, wie SAXS und CryoEM, in einem definierten Beladungszustand einzusetzen. Die mehrfach-geladenen Ionenmassen für die Variationen der Beladungszustände wie die Apo- und Holo-Form mit Thiol- und Amino-Terminus sowie zusätzlich verknüpfte Derivate müssen bei multimodularen Zielproteinen für jedes PCP-Fragment berechnet und anschließend kontrolliert werden (Abb. 44).

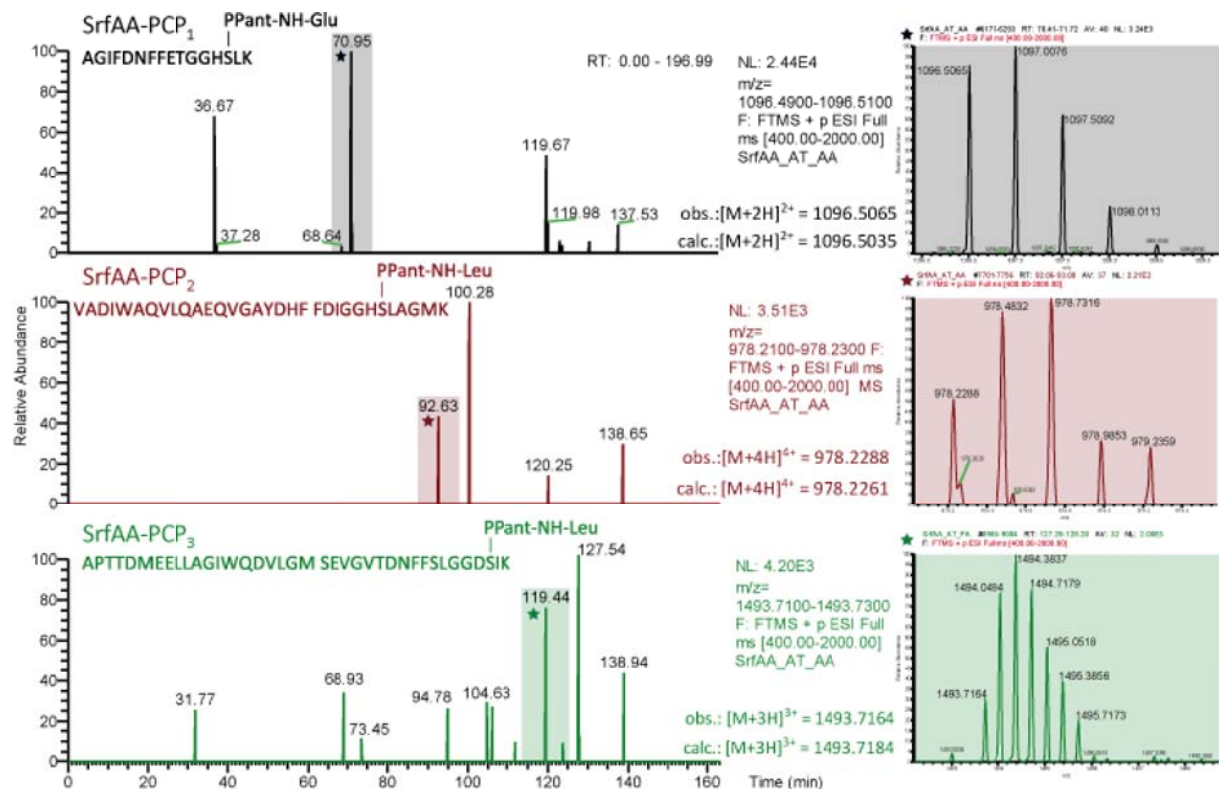


Abbildung 44: Die SrFAA-PCP-Fragmente im beladenen Zustand, sowie die Spektren der entsprechenden Ionenmassen ([M+2H]<sup>2+</sup> bei PCP<sub>1</sub>, [M+4H]<sup>4+</sup> bei PCP<sub>2</sub> und [M+3H]<sup>3+</sup> bei PCP<sub>3</sub>).

Nur wenn alle möglichen PCP-Fragmente ausschliesslich im gewünschten Beladungszustand identifiziert werden können, wurde die Probe für die weitergehenden Untersuchungen benutzt. Bei einer Beladung mit Amino-CoA und der anschliessenden Verknüpfung mit der korrespondierenden Aminosäure, sowie einem Massenbereich von [M+H]<sup>+</sup> bis [M+5H]<sup>5+</sup>, müssen für die einzelnen möglichen Beladungszustände einer trimodularen nicht-ribosomalen Peptidsynthetase, Apo-, Holo-S- (SH-Terminus), Holo-NH- (NH<sub>2</sub>-Terminus) und Holo-NH-AA-Form (Amino-Terminus mit der Verknüpften Aminosäure), 60 mögliche

Ionenmassen überprüft werden, damit die spätere Datenintegrität gewährleistet ist.

## 5.5 SAXS mit modularen NRPS

Die Untersuchungen mittels Kleinwinkelstreuung ermöglichen Einblicke in die strukturelle Orientierung dynamischer Domänen in Lösung. In erster Linie sollte durch die Untersuchung von SrfAC das Modell der *Workbench* unter nativen Bedingungen bestätigt werden. Zusätzlich sollte das Akzeptor-Modell (*acceptor site only*) bestätigt, oder um dynamische Möglichkeiten in der Reorganisation der Domänen erweitert werden. Alle verwendeten Proteine wurden durch Affinitätschromatographie gereinigt und anschliessend mittels Gelfiltration von unerwünschten Ionen bzw. Proteinen befreit. Anschließend wurden die Proben auf eine definierte Konzentration ( $> 6 \text{ mg/mL}$ ) ankonzentriert, der dabei entstandene Durchlauf wurde als Gegenpuffer und zur Herstellung der Verdünnungsreihen verwendet.

### 5.5.1 SrfAC – *Workbench* Untersuchung mittels SAXS

Der Vergleich der SAXS-Streukurven zeigt eine deutliche Abweichung zwischen der berechneten SrfAC-S1003A- (Abb. 45, grün) und gemessenen SrfAC-Streukurve (Apo-PCP; rot). Der Wert für die Anpassungsqualität zwischen der berechneten und gemessenen Streukurve zeigt einen erhöhten Wert ( $\chi^2 = 7.2$ ), der deutlich über dem Schwellenwert von

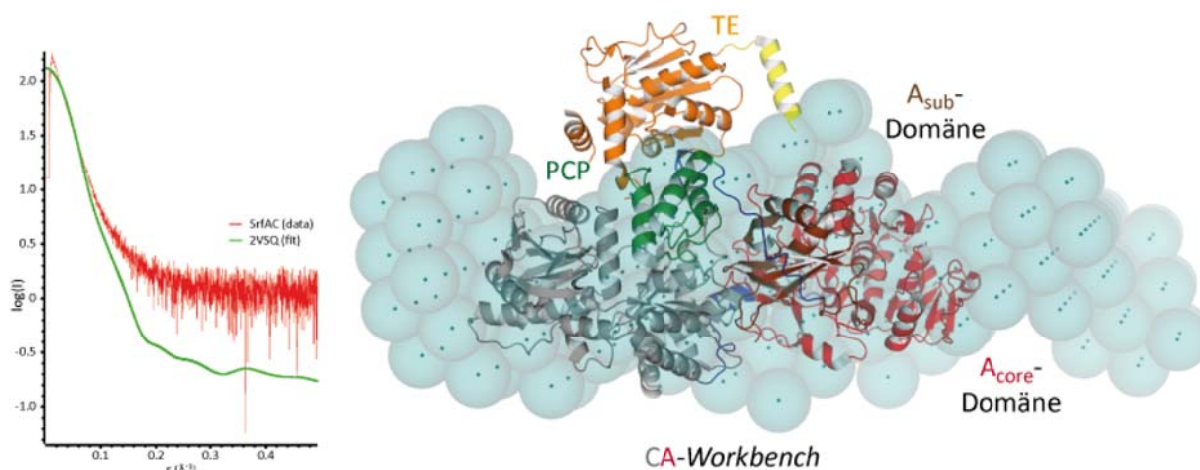


Abbildung 45: SrfAC Kleinwinkelstreucurve (rot) im Vergleich mit dem Strukturmodell (grün). Die SrfAC-Domänen (pdb: 2VSQ) wurden wie folgt eingefärbt: C-Domäne in grau,  $A_{core}$ -Domäne in rot,  $A_{sub}$ -Domäne in braun, PCP in grün und TE-Domäne in orange (myc-Helix in gelb). Ein grosser Teil der TE-Domäne befindet sich außerhalb der Elektronendichte des ab initio-Modells (cyan), während Teile der Elektronendichte nicht besetzt werden können.

1.5 liegt. Vergleicht man zusätzlich das SrfAC-S1003A-Strukturmodell mit dem nach den Messdaten erstellten *ab initio*-Modell (Abb. 45; cyan), wird die abweichende Deckung der beiden Modelle sichtbar. Das *ab initio*-Modell zeigt eine elongierte Elektronendichte, die mit dem kompakten SrfAC-S1003-Strukturmodell nicht zu struktureller Deckung zu bringen ist. Um eine bessere Deckung zwischen dem berechneten Modell und den SrfAC-Werten erreichen zu können, müssen die PCP- und TE-Domäne dissoziiert vorliegen. Damit die Abweichung zwischen dem *ab initio*-Modell und dem SASREF *rigid body modeling*-Resultat verringert wird, muss die sequenzbedingte maximale Linkerlänge zwischen den Domänen geändert werden. Dadurch erhöhen sich die Freiheitsgrade der einzelnen Domänen womit die Auswahl der relativen Orientierungen zueinander zunehmen. Durch diese Vorgehensweise kann eine bessere Übereinstimmung mit dem *ab initio*-Modell erreicht werden. Setzt man die maximale Länge der Linker auf 50 Å kann das SrfAC-Modell für den dissoziierten Zustand berechnet werden (Abb. 46).

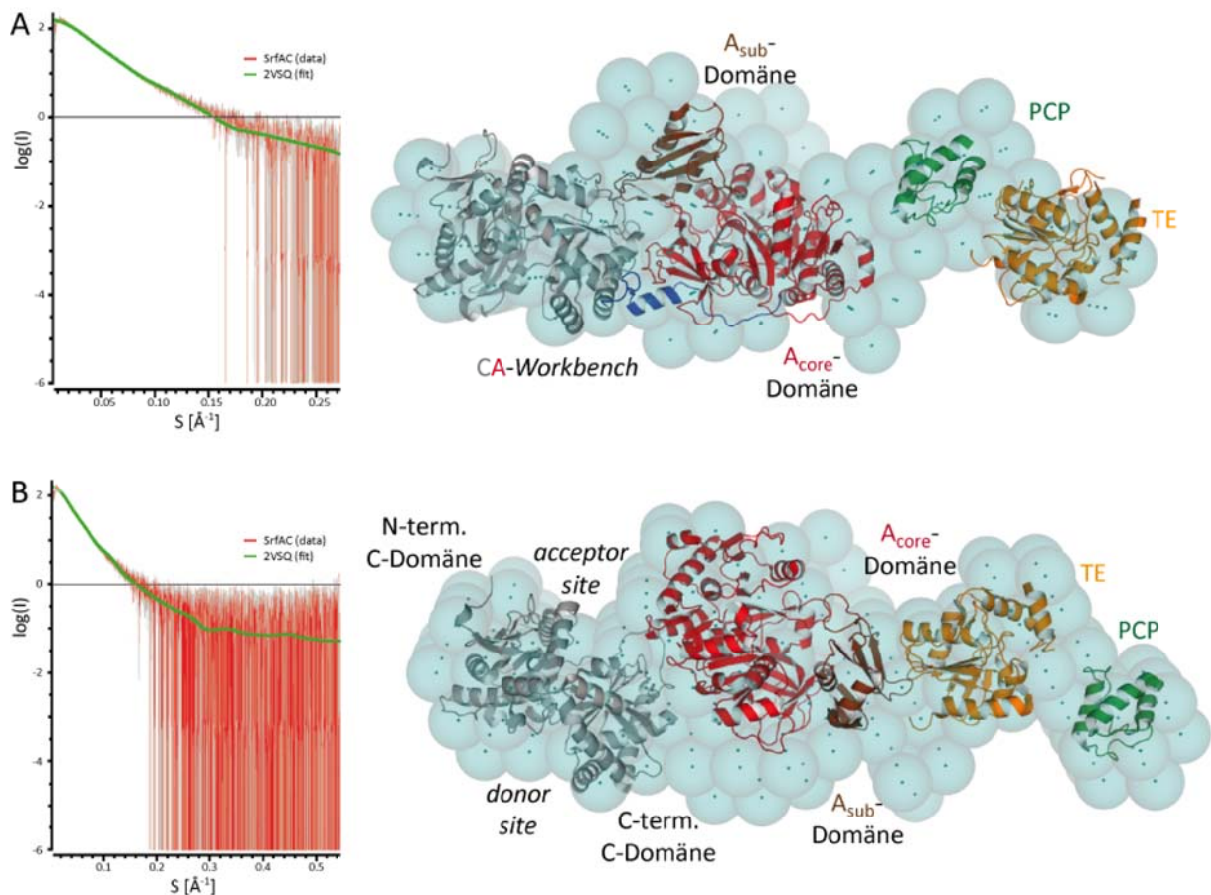


Abbildung 46: Kleinwinkelstreuksurve (rot) im Vergleich mit dem Konformationsmodell (grün). (A) Berechnetes *ab initio*-Modell (cyan) im Vergleich mit den Strukturmodellen sowie den dazugehörigen Streuksurven für festes C-A-Interface (*Workbench*). (B) Die dynamische C- und A-Domänen Interaktion.



Mit einem festen Interface zwischen der C- und A-Domäne und den dissoziierten PCP- und TE-Domänen werden die Freiheitsgrade der dynamischen Domänen erhöht und so eine bessere Anpassungsqualität erreicht. Für das feste C-A-Interface zeigt das Strukturmodell einen Wert für die Anpassungsqualität von  $\chi^2 = 1.64$  (Abb. 46A), wenn alle Domänen die maximalen Freiheitsgrade besitzen einen Wert von  $\chi^2 = 1.24$  (Abb. 46B). Auch wenn der Wert für die maximalen Freiheitsgrade erwartungsgemäß eine bessere Anpassungsqualität zeigt, können durch die gemessenen Daten beide Modelle bestätigt werden. Bei dem vollständig dissoziierten Modell ist die Akzeptorseite der C-Domäne durch die A-Domäne blockiert und ist nicht durch die über 150 Å entfernte PCP-Domäne erreichbar, während bei dem *Workbench*-Modell die Funktionalität des Moduls völlig erhalten ist.

### 5.5.2 Kleinwinkelstreuung an dimodularen NRPS (TycB<sub>2-3</sub>)

Durch die Erhöhung der Freiheitsgrade im SrfAC-Modell wird eine höhere Übereinstimmung mit den Messwerten erreicht. Daher bietet die Messung an dimodularen NRPS-Systemen eine wesentlich größere Herausforderung, da hier die erhöhte Anzahl an Domänen gleichbedeutend ist mit einer wesentlich höheren Anzahl an deren Orientierungsmöglichkeiten. Dabei steigt die Anzahl der Freiheitsgrade sowie deren Kombinationen mit jedem Modul exponentiell an. Um diesem Effekt entgegenzuwirken wurden mehrere Beladungszustände sowie Domänen-Kombinationen vermessen. Für die verschiedenen Domänen-Kombinationen wurden die primären Parameter ermittelt (Abb. 47). Um Abweichungen von erwarteten linearen oder konstanten Verlauf wurden die primären Parameter gegen die gemessenen Konzentration aufgetragen.

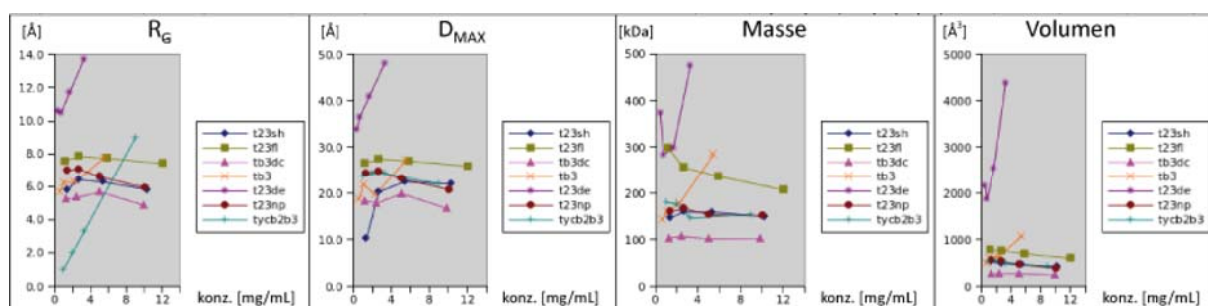


Abbildung 47: Aus den Streukurven ermittelte primäre Parameter: Gyrationradius  $R_G$  (Å), maximale Partikelgröße  $D_{MAX}$  (Å), Masse (kDa) und Volumen (Å<sup>3</sup>).

Sofern eine Differenz von dem linearen oder konstanten Verlauf festgestellt wurde, konnte

der abweichende Wert aus der Datenreihe entfernt werden. Der Gyrationradius ( $R_G$ ) und die maximale Distanz der an dem Streubeitrag beteiligten Partikel ( $D_{MAX}$ ) können direkt aus den Streukurven abgeleitet werden, das Volumen und die Masse werden durch Vergleichsmessungen mit einem Kontrollprotein, in diesem Fall BSA, berechnet. Der Datensatz t23de (Apo-TycB<sub>2-3ΔE</sub> [ATCAT]) zeigte ein hohes Hintergrundrauschen was die Berechnung von zuverlässigen Werten für die P(r)-Funktion und somit  $D_{MAX}$ -Werten unmöglich machte. Daher wurden die Daten für dieses Zielprotein bei allen späteren Betrachtungen nicht hinzugezogen. Für den Datensatz tycb3 (Apo-TycB<sub>3</sub> [CATE]) konnte durch die Entfernung einzelner abweichender Werte die Kohärenz des gemessenen Datensatzes erhalten werden. Bei steigender Konzentration kann hier ein Anstieg für die  $R_G$ - und  $D_{MAX}$ -Werte verzeichnet werden, was durch die Agglomeration bei hoher Konzentration verursacht wird. Nach der Bestätigung durch die DLS-Messungen wurden bei weiteren Berechnungen die Werte für die höchste Konzentration ausgeschlossen (Abb. 48). Die DLS-Datensätze ermöglichten eine nach der Partikelgröße aufgelöste Analyse der Probe. Dadurch konnten die zu der Partikelgröße proportionalen Streubeiträge der agglomerierten Proteine sowie anderer Fremdpartikel ermittelt und somit ihr Beitrag an dem berechneten *ab initio*-Modell minimiert werden. Für die drei unterschiedlichen Domänen-Kombinationen, TycB<sub>2-3FL</sub>, TycB<sub>2-3</sub> und TycB<sub>3</sub> wurden die Werte für die Partikelgröße sowie die berechnete Masse ermittelt (Tab. 17).

Zielprotein	TycB <sub>2-3</sub> FL-NH-Phe	Apo-TycB <sub>2-3FL</sub>	TycB <sub>2-3</sub> -SH	TycB <sub>2-3</sub> -NH-Phe	Apo-TycB <sub>2-3</sub>	Apo-TycB <sub>3</sub>
Datensatz	t23flnp	t23fl	t23sh	t23np	tycb2b3	tb3
Domänenorganisation ( $M_w$ )	[CATCATE] (285 kDa)	[CATCATE] (285 kDa)	[ATCATE] (218 kDa)	[ATCATE] (218 kDa)	[ATCATE] (218 kDa)	[CATE] (169 kDa)
$D_{MAX} - M_{w(calc)}$	37.5 – 257	32.5 – 250	27.0 – 154	30.5 – 158	28.5 – 170	18.0 – 160

Tabelle 17: Die SAXS-Datensätze der Zielproteine und deren primäre Parameter.

Für gemessene Parameter, insbesondere durch die berechnete Masse, können Abweichungen zwischen den berechneten und tatsächlichen Parameter der Zielproteine festgestellt werden. Während die Abweichung zwischen dem Molekulargewicht und der berechneten Masse durch die flexible Domänenorganisation und dem dadurch begründeten verringerten mittleren Volumen begründet ist, zeigt die Differenz im maximalen Partikel-

abstand ( $D_{MAX}$ ) eine eindeutige Abhängigkeit von dem Beladungszustand der PCP-Domäne. Dabei beinhalten TycB<sub>2-3FL</sub> ([CATCATE]) und TycB<sub>2-3</sub> ([ATCATE]) zwei PCP-Domänen, jedoch besitzt die entsprechende PCP-Domäne in TycB<sub>2-3</sub> aufgrund der fehlenden TycB<sub>2</sub>-C-Domäne eine geringere Anzahl Interaktionspartner. Generell zeigen alle *ab initio*-Modelle der gemessenen NRPS-Systeme eine Abweichung zu den strukturellen Modellen (CATCATE und ATCATE). Dabei ist eine deutliche Ausdehnung entlang der Längsachse zu verzeichnen, die bei einer erhöhten Modulanzahl deutlicher wird.

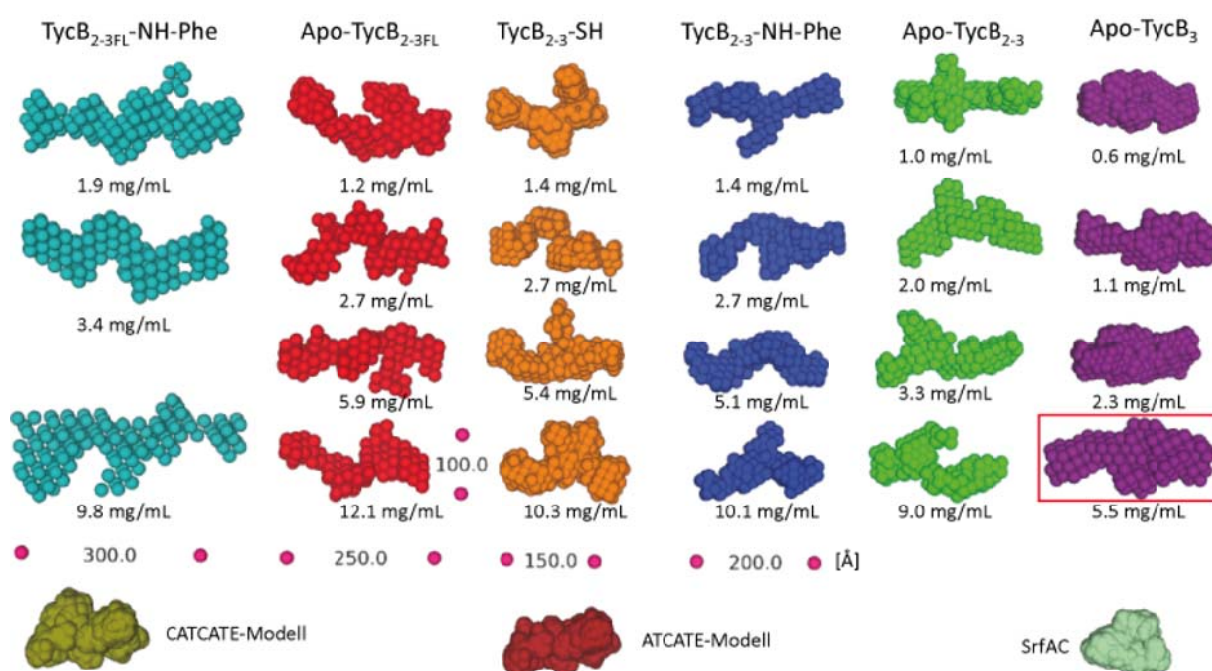


Abbildung 48: Die *ab initio*-Modelle für den Konzentrationsverlauf der gemessenen Zielproteine, sowie im Vergleich die Sphärendarstellung der NRPS-Modelle und SrfAC. Wegen der zu hohen Abweichung wurden die Werte für TycB<sub>3</sub> bei 5.5 mg/mL ausgeschlossen.

Im Strukturmodell interagiert die Elongationsseite der PCP-Domäne mit der Akzeptorseite und die Transferseite mit der Donorseite der entsprechenden C-Domäne. Im Akzeptor-Modell ist die aktive Seite, die für die kovalente Bindung des PPant-Armes verantwortlich ist, der Akzeptor-Seite der sich im gleichen Modul befindlichen C-Domäne zugewandt. Eine Auflösung dieser Interaktion unter nativen Bedingungen, wie bereits für SrfAC durch die PCP- und TE-Domänen-Dissoziation gezeigt worden ist, resultiert zwangsläufig in der Erhöhung der Freiheitsgrade der einzelnen Module zueinander. Dadurch nimmt die dimodulare Struktur, anders als im Modell, eine länglichere globale Struktur an. Die zusätzliche Beladung der PCP-

Domäne mit dem PPant-Arm erzeugt innerhalb der einzelnen Module eine kompakte Struktur, die Strukturkompartimente, die an dieser Interaktion nicht beteiligt sind zeigen sich dennoch durch eine hohe Flexibilität aus. Das Zielprotein TycB<sub>2-3</sub> [ATCATE] nimmt für die drei verschiedenen Beladungszustände zwei globale Orientierungen ein (Abb. 49A). Während das Zielprotein TycB<sub>2-3</sub>-NH-Phe ( $D_{MAX} = 30.5$ ) eine höhere Partikelgröße als das Apo-Zielprotein ( $D_{MAX} = 28.5$ ) aufweist, zeigen beide globale Strukturen eine gute Übereinstimmung in ihrer Form. Dagegen zeigt das Holo-TycB<sub>2-3</sub>-Zielprotein eine signifikante Abweichung in der Form und der Partikelgröße (Abb. 49A). Während das Holo-TycB<sub>2-3</sub>-Zielprotein eine kompakte sphärische Form aufweist ( $D_{MAX} = 27.0$ ), zeigen TycB<sub>2-3</sub>-NH-Phe und die Apo-Form eine längliche Form. Mit der N-terminalen C-Domäne zeigen die Zielproteine TycB<sub>2-3FL</sub>-NH-Phe und Apo-TycB<sub>2-3FL</sub> deutlich signifikantere Unterschiede zueinander (Abb. 49B). Das Zielprotein TycB<sub>2-3FL</sub>-NH-Phe zeigt wesentlich höhere  $R_g$ - und  $D_{MAX}$ -Werte als das Apo-Zielprotein, sowie eine rigide jedoch gestreckte globale Struktur.

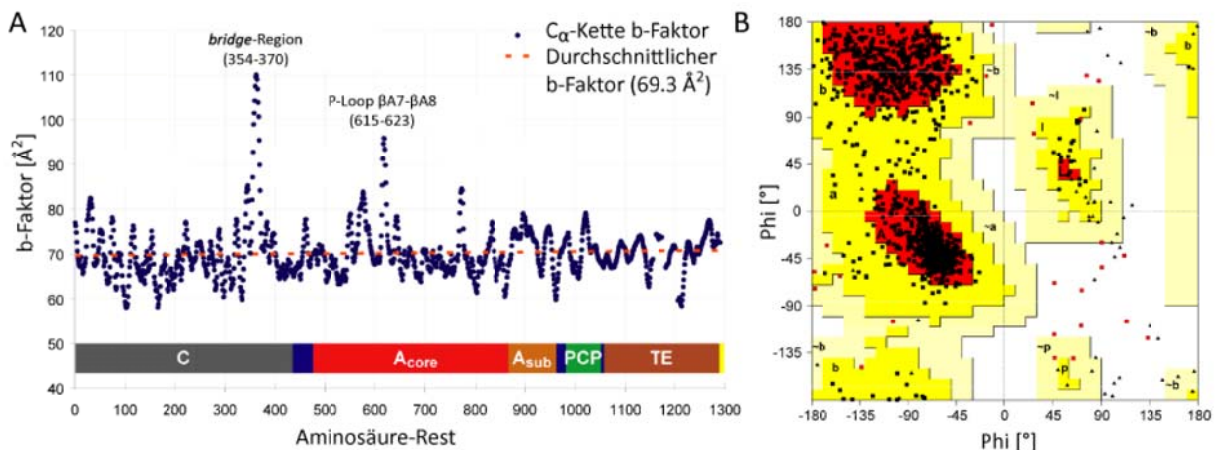


Abbildung 49: Veränderungen an der globalen Struktur der dimodularen NRPS-Systeme initiiert durch die verschiedenen Beladungszustände der dynamischen PCP-Domäne. (A) Die Streukurven nach Abzug des Puffers für TycB<sub>2-3</sub> [ATCATE] ohne und mit dem PPant-Arm, sowie mit dem Amino-Derivat und gebundener Aminosäure. Die Form der so erstellten Streukurve ist maßgeblich für die spätere Form der errechneten ab initio-Modelle verantwortlich. Der t23sh-Datensatz des Holo-TycB<sub>2-3</sub>-Zielproteins (CoA-SH) zeigt eine deutliche Abweichung von dem tycb2b3- und t23np-Datensatz, was sich in der späteren abweichenden globalen Struktur bemerkbar macht. (B) Kratky-Plot für Holo- und Apo-TycB<sub>2-3FL</sub> zeigt keine signifikante Auswirkungen auf die globale Domänenanordnung, doch einen höheren  $D_{MAX}$ -Wert für den t23flnp-Datensatz.

Obwohl die Änderung des Beladungszustandes der PCP-Domänen die dynamische Veränderungen der globalen Struktur durch die Reduktion der Freiheitsgrade einschränkt, zeichnen sich die gemessenen Zielproteine immer noch durch ein hohes Maß an lokalen



dynamischen Konformationsänderungen aus. Durch die permanente Änderung kann das gemessene Mittel aller Konformationen zu keinem signifikanten Resultat führen. Die *ab initio*-Modelle zeigen im Ansatz eine schraubenförmige Struktur, doch die Erstellung eines zufriedenstellenden Strukturmodells für die entsprechenden Domänen in der vorliegenden Modellelektronendichte ist nicht möglich.

## 6 Diskussion

Untersuchungen der Architektur multimodularer Enzymkomplexe zeigen sowohl ihre Komplexität und strukturelle Vielfalt als auch die funktionelle Abhängigkeit von kleinen, prothetische Gruppen bindenden Domänen, den *swinging domains*. Für das bessere Verständnis dieser multimodularen Komplexe sind nicht nur die strukturellen Daten über die dynamischen Domänen sondern insbesondere auch die flankierenden Linkerbereiche von großer Wichtigkeit. In vielen Fällen werden die dynamischen Domänen ohne die flankierenden Linkerbereiche durch kristallographische Methoden oder NMR-Experimente separat als einzelne Proteine strukturell untersucht. In den bisher bekannten globalen Strukturen der fFAS[54] oder mFAS[49] sowie der PKS[58, 70] ist die Elektronendichte der dynamischen ACP-Domäne sowie der umgebenden Linkerbereiche nicht aufgelöst. Nachfolgende Untersuchungen beschäftigen sich mit der Interaktion dieser dynamischen Domänen und benutzen die Daten der multimodularen Kristallstrukturen als strukturelle Plattform[72, 237]. Desweiteren ermöglicht die Kenntnis über die strukturelle Anordnung und Interaktionen aller Domänen eines NRPS-Moduls einen detaillierten Einblick in die Organisation dieser Systeme und kann als Templat für die Erstellung von Modellen anderer verwandter Systeme dienen. Dabei kann das räumliche Strukturmodell durch die Angabe der Sequenzen strukturell- oder funktionsverwandter Proteine anhand des vorgegebenen Templates erstellt werden (Tab. 18).

Bezeichnung	DEBS <sub>3</sub>	DEBS <sub>5</sub>	fFAS ( <i>Wheel+Dome</i> )	fFAS	fFAS	mFAS	SrfAC
pdb-Code	2QO3	2HG4	2UVB + 2UVC	3HMJ	2UV8	2VZ9	2VSQ
Auflösung	2.59 Å	2.73 Å	3.10 Å	4.00 Å	3.10 Å	3.30 Å	2.60 Å
Organismus	<i>S. erythraea</i>		<i>T. lanuginosus</i>	<i>S. cerevisiae</i>		<i>Sus scrofa</i>	<i>B. subtilis</i>
PMP-Templat	287	553	9	249	1326	2301	5957

Tabelle 18: Die Verwendung von Strukturmodellen multimodularer Enzymkomplexe als PMP-Templat (*Protein Model Portal*-Templat; 12.2011)

Die Anzahl der Modelle, ausgehend von der SrfAC-Struktur als PMP-Templat[238], gibt ihren Stellenwert für das Verständnis der NRPS-Systeme wieder. Dies liegt in der hohen Verbreitung der NRPS-Systeme für die Biosynthese von Sekundär-Metaboliten, als auch in

den besonderen und bislang einzigartigen Merkmalen der SrfAC-Struktur begründet. Neben der strukturell definierten PCP-Domäne, mitsamt den flankierenden Linkerregionen und den verschiedenen Interaktionsflächen der Domänen, zeigt die SrfAC-Struktur die bislang erste Struktur einer  $^D C_L$ -Kondensationsdomäne (katalysiert die Kondensationsreaktion zwischen einer  $D$ -Aminosäure von der Donorseite mit einer  $L$ -Aminosäure von der Akzeptorseite), das Zusammenspiel der C- und A-Domäne in der Bildung der katalytischen *Workbench* sowie das COM-Helix-Hand-Interaktionsprinzip zwischen den einzelnen nichtribosomalen Peptidsynthetasen. Diese einzigartigen Merkmale dienen einem besseren Verständniss der NRPS-Systeme.

## 6.1 NRPS Linker- und Intercore-Bereiche mit SrfAC als Templat

In dem Strukturmodell des SrfAC-Terminationsmoduls sind der C-A- sowie der A-PCP-Linkerbereich in der Elektronendichte aufgelöst. Während der C-A-Linker durch starke Interaktionen in die *Workbench* integriert ist, zeigt der A-PCP-Linker keine signifikanten Interaktionen mit den anderen Bereichen der SrfAC-Struktur. Für die systematische Analyse der Linkerregionen und der Intercore-Bereiche (Bereiche, die zwischen den hoch konservierten Core-Motiven[91, 97, 100, 116, 239] lokalisiert sind) wurden die Sequenzen verschiedener NRPS-Systeme[240-263], von Anabaenopeptin[264] bis Viomycin[265] Biosynthese-Modulen, selektiert und gespeichert. Zusätzlich wurden die erfassten Module in Elongations-, Terminations- und COM-Module klassifiziert.

Die Klasse der Elongationsmodule (42 Sequenzen) enthält die Sequenzen zwischen dem Core-Motiv der vorausgehenden Donor-PCP-Domäne und dem ersten Core-Motiv der nachfolgenden  $^L C_L$ -Kondensationsdomäne ( $T1[GGHSL]^{n-1} - C1[SxAQxR(LM)(WY)xL]^{n+1}$ ). Dadurch können gezielt Informationen über die modolverbindenden Linkerbereiche berechnet werden. Die Klasse der COM-Module (21 Sequenzen) beinhaltet Module die an der Interaktion zweier Synthetasen beteiligt ( $N\text{-Terminus}^n - C1[SxAQxR(LM)(WY)xL]^{n+1}$ ). Da der Interaktionspartner die COM-Helix-tragende[133] Epimerisierungsdomäne enthält, ist die Kondensationsdomäne des COM-Moduls stets eine  $^D C_L$ -Domäne. Die Klasse der Terminationsmodule (23 Sequenzen) definiert Module die am C-Terminus der letzten Synthetase lokalisiert und katalysieren durch die Thioesterase-Domäne die Freisetzung des

Produktes ( $C1[SxAQxR(LM)(WY)xL]^n - C\text{-Terminus}^n$ ). Als Strukturtemplat wurde für die COM- sowie Terminationsmodul-Sequenzen das SrfAC-Strukturmodell verwendet (Abb. 50). Durch die Kombination mit dem  $TycC_{5-6}$ -PCP-C-Strukturmodell[121] konnte mit dem Programm Modeller v9.0[266-267] ein Strukturmodell für das  $TycC_5$ -Elongationsmodul ( $T1[GGHSL]^{TycC4} - C1[SxAQxR(LM)(WY)xL]^{TycC6}$ ) erstellt werden (Abb. 50). Die Sequenzidentität betrug dabei 29% und der RMSD-Wert lag bei 1.6 für 1056  $C_\alpha$ -Atome. Die so klassifizierten Sequenzen wurden durch Alignment-Software (siehe Kapitel 4.7) ausgerichtet und in ALINE[268] weiterverarbeitet, so dass die Verteilung der konservierten Reste anhand der Strukturmodelle in PyMOL[230] visualisiert werden konnte (Abb. 50).

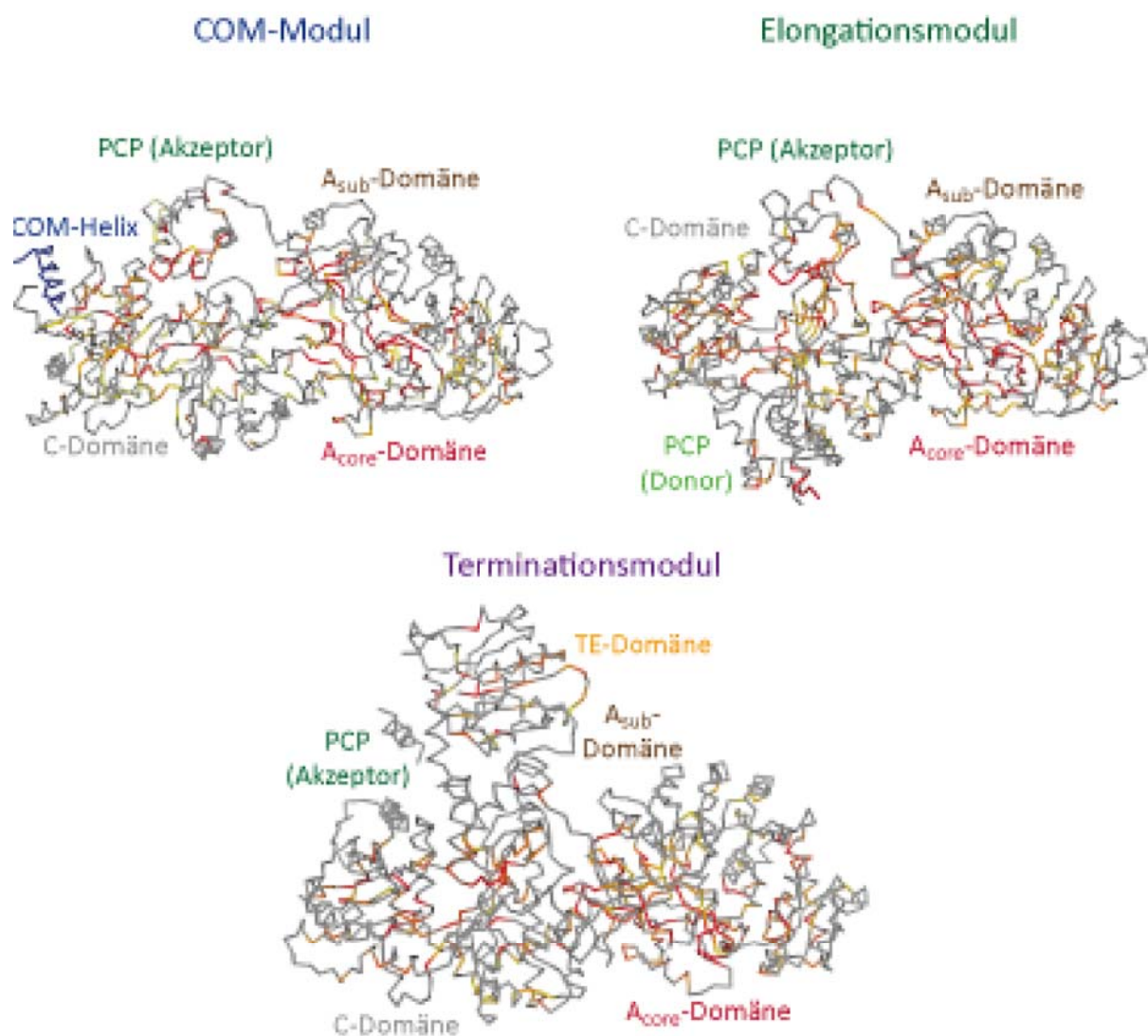
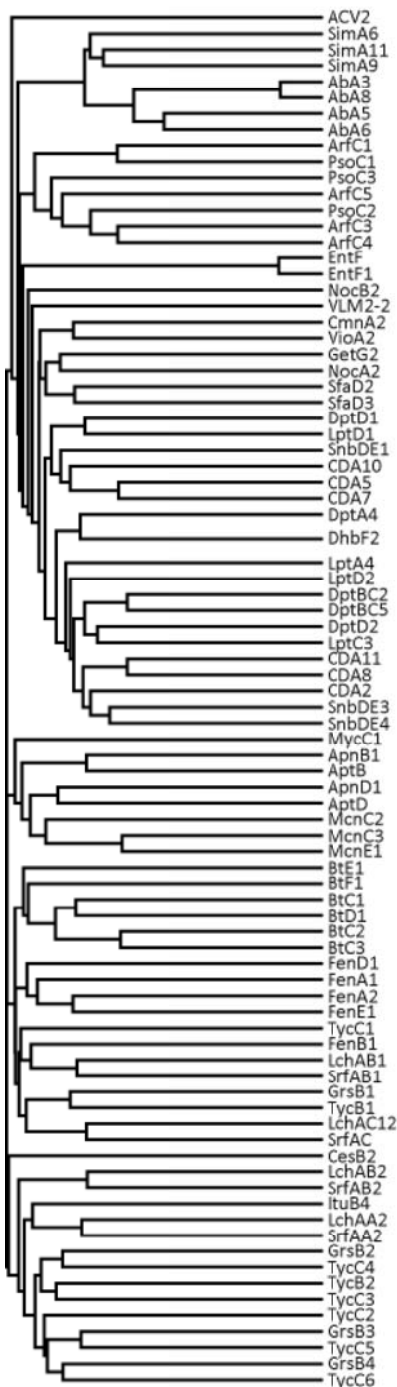


Abbildung 50: Verteilung der konservierten Reste für die erfassten NRPS-Module. Die einzelnen Reste wurden ihrer Sequenzhomologie entsprechend eingefärbt, 60%-80% Übereinstimmung von gelb nach orange, 80%-100% Übereinstimmung von orange nach rot.

Da bereits sichergestellt wurde, dass die einzelnen Sequenzen in dem erstellten Alignment bereinstimmen, so dass die Core-Motive als die signifikanten Marker benutzt und die festgestellten Abweichungen nachvollziehbar erklärt werden können (Kapitel 8.4), ist eine genaue Berechnung der Länge sowie der Aminosäurenverteilung anhand der Merge-Klasse, in der alle Sequenzen vereint wurden, möglich.

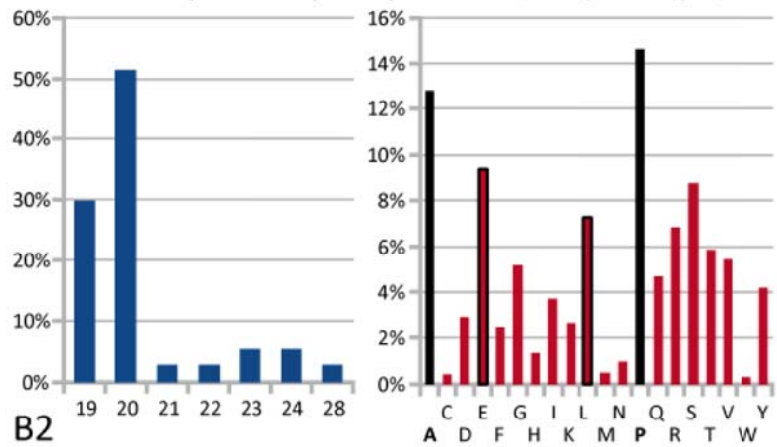
Bedingt durch ihre Funktion enthalten nur die Elongationsmodule den PCP-C-Linkerbereich, wodurch die Anzahl der berücksichtigten Sequenzen für die Berechnung von diesem Bereich bei 40 liegt. Die Berechnung des PCP-C Linkerbereichs zeigt einen Mittelwert von 20.4 ( $\sigma = 1.8$  Aminosäuren, 19 in TycC<sub>5</sub> PCP-C), die des C-A-Linkerbereichs einen von 33.7 ( $\sigma = 1.0$  Aminosäuren, 33 in SrfAC) und die Berechnung des A-PCP-Bereichs einen Mittelwert von 18.4 ( $\sigma = 1.5$  Aminosäuren, 19 in SrfAC). Die niedrige Abweichung zeigt, dass die jeweilige mittlere Linkerlänge eindeutig bevorzugt wird, obwohl kürzere Versionen ebenfalls signifikant vertreten sind (Abb. 50). Auch wenn der Mittelwert des A-PCP-Linkerbereichs eine Linkerlänge von 18 Aminosäuren voraussagt, zeigt das nach Linkerlängen aufgeschlüsselte Histogramm eine bevorzugte Länge von 19 bzw. 17 Aminosäuren zwischen dem letztem Motiv der A-Domäne (A10) und der ersten  $\alpha$ -Helix der PCP-Domäne ( $\alpha$ P1). Die aus der Kleinwinkelstreuung (SAXS) resultierenden Modelle mit solvatisierten Atommodellen für das SrfAC bestätigen nicht nur das Bestehen der *Workbench* im solvatisierten Zustand, sondern zeigen auch indirekt die Unterschiede in der Natur des C-A- und des A-PCP-Linkers. Der 33 Aminosäuren lange L-förmige C-A-Linker ist an den vielzähligen Wechselwirkungen zwischen der C- und der A-Domäne beteiligt. Der Linker verbindet den 50 Å auseinander liegenden C-Terminus der C-Domäne mit dem N-Terminus der A-Domäne, die Interaktionen bleiben im SAXS-Modell bestehen. Der dagegen nur 19 Aminosäuren lange, A-PCP-Linker zeigt keine nennenswerten Interaktionen mit dem Rest der Struktur und verbindet den 44 Å auseinander entfernten C-Terminus der A-Domäne mit dem N-Terminus der PCP-Domäne. Seine flexible Struktur ermöglicht der PCP-Domäne die erforderliche Bewegungsfreiheiten zu den korrespondierenden Katalysezentren zu gelangen, ohne die Verkürzung der Reichweite durch die Ausbildung sekundärer Strukturelemente. Im SAXS-Modell dissoziiert die PCP-Domäne von der *Workbench*, ist jedoch nicht diffus, sondern eindeutig positioniert, was in der Natur des flexiblen A-PCP-Linkers begründet liegt.

## A Kladogramm-Baum



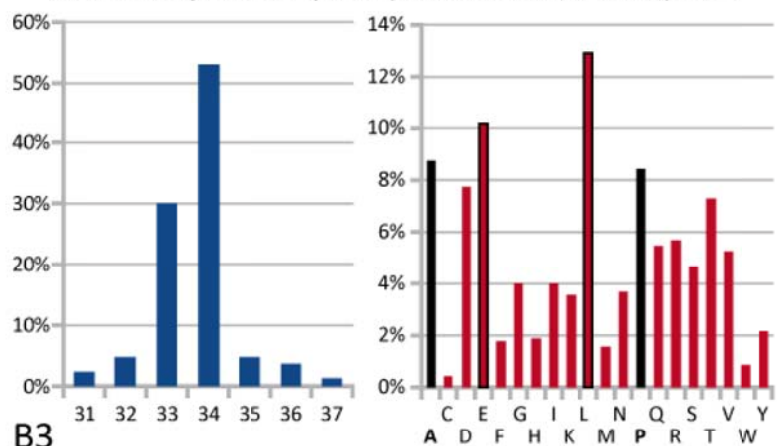
## B1 Analyse der Linkerlänge und der AA-Verteilung

PCP-C Linker-Analyse aus 40 Sequenzen [AAHIAE - SFAQKRMV ( 58 - 134 )]: 20,4 +/- 1,8



## B2

C-A-Linker-Analyse aus 82 Sequenzen [QQLIQ-N - PLTYW ( 720 - 777 )]: 33,7 +/- 1,0



## B3

A-PCP-Linker-Analyse aus 82 Sequenzen [VNKR - NEMEET (1408 - 1434)]: 18,4 +/- 1,5

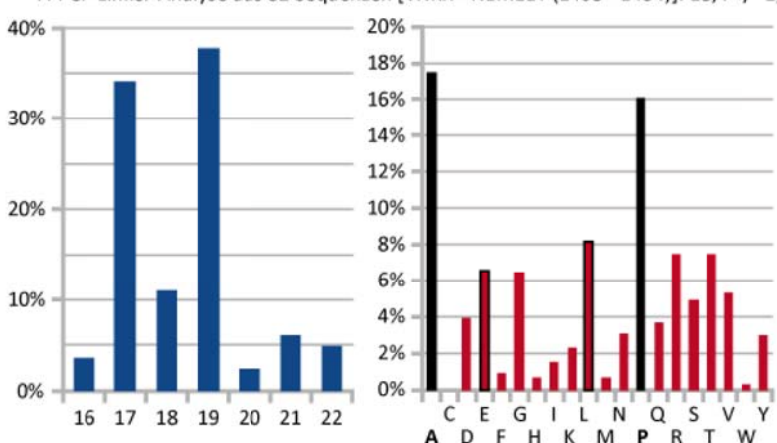


Abbildung 51: Analyse der Linkerbereiche aufgelöst nach Länge und Verteilung. (A) Kladogramm der 86 verwendeten NRPS-Sequenzen, davon 21 Module mit dem COM-Hand-Motiv [CAT-], 42 Elongationsmodule [-TCAT-] und 23 Terminationsmodule [-CATTE]. (B) Analyse der Linkerbereiche, aufgelöst nach Länge (blau; Achsen: Länge (x), Häufigkeit (y)) und Verteilung (rot; Achsen: Aminosäure (x), Häufigkeit (y)). Die Linker mit einer bestimmten Länge scheinen bevorzugt, auch wenn um eine bzw. zwei Aminosäuren kürzere Versionen ebenfalls signifikant vertreten sind. Dabei zeigen die die PCP flankierenden Linker höhere Werte für Prolin und Alanin, was auf eine, wie bereits in SrfAC vorgefundene, flexible Linkerstruktur hindeutet (B1 und B3). Der an den C-A-Interaktionen beteiligte Linker zeigt dagegen eine andere Verteilung der Aminosäuren (B3).

Die flexible Struktur der Linker wird durch einen hohen Anteil (ca. 30%) an Prolin- und Alanin-Resten in den die PCP-Domäne flankierenden Linkerbereichen erreicht (Abb. 51). Die Analyse der Linkerbereiche zeigt daher Analogien zu den anderen *swinging domains* beinhaltenden Systemen, wie PDH[9, 269], BCCP[270], PKS[271] oder FAS[272], in denen auch der hohe Anteil an Prolin- und Alanin-Resten in den Linkerbereichen die Ausbildung von Sekundärstrukturelementen verhindert.

Das TycC<sub>5</sub>[121]- und das SrfAC[236]-Strukturmodell beschreiben durch die PCP-C- und A-PCP-Linkerbereiche, die sich durch niedrige b-Werte, geringe Interaktionen mit dem Rest des Strukturmodells sowie durch vollständig aufgelöste Elektronendichte auszeichnen, den ersten vollständig strukturell aufgeklärten Linkerbereich einer *swinging domain* (PCP-Domäne).

### 6.2 SrfAC Strukturmodell als NRPS-Strukturtemplat

Das SrfAC-Strukturmodell ermöglicht durch seine Domänenkonstellation einen Einblick in die Funktionsweise der NRPS-Module. Neben der Thioesterase (TE-Domäne), die nur in den terminalen Modulen vorkommt, beinhaltet das SrfAC-Strukturmodell die C-, A- und die PCP-Domäne, durch die ein Elongationsmodul definiert wird. Das Strukturmodell der SrfAC Thioesterase in der offenen Konformation, die für die alleinstehende Thioesterase bereits publiziert[124] wurde, zeigt durch fehlende Kristallkontakte in dem SrfAC-Strukturmodell keine aufgelöste Elektronendichte für die Loop-Bereiche der Lid-Region. Durch die strukturelle Untersuchung des Bidomänen-Systems PCP-TE aus EntF[125] konnte mit der  $\alpha$ TE1 die Interaktionsfläche zwischen der im AH-Zustand arretierten PCP-Domäne und der iterativ arbeitenden Thioesterase identifiziert werden (Abb. 52). Dabei liegt die Alanin<sup>48</sup> – Serin<sup>180</sup> Distanz mit 19 Å in der Reichweite des PPant-Arms. Da die PCP-Domäne in dem SrfAC Strukturmodell mit der C-Domäne interagiert, ist sie im Vergleich zu der EntF PCP-Domäne von der Interaktionsfläche der TE-Domäne abgewandt, die Alanin<sup>1003</sup> – Serin<sup>1120</sup> Distanz beträgt dabei 44 Å. Die  $\alpha$ TII-Helix der PCP-Domäne, die das konservierte Serin bzw. die Alanin Mutation (Alanin<sup>1003</sup> in SrfAC und Alanin<sup>48</sup> in EntF) trägt, ist um ca. 180° gedreht. Durch die Interaktion mit der PCP-Domäne richtet sich die  $\alpha$ TEIV-Helix um ca. 45° neu aus. Der in SrfAC nicht aufgelöste  $\alpha$ TEIV- $\alpha$ TEV-Loop sitzt durch Umoorientierung der  $\alpha$ TEIV-Helix über dem



PPant-Kanal und kann so den Zugang zum Kanal, durch Öffnen bzw. Schließen, dynamisch regulieren[124]. Neben der Thioesterase bilden die weiteren essenziellen SrfAC-Domänen die strukturelle Vorlage für NRPS-Systeme. Obwohl die Funktion, wie auch die Faltung dieser Domänen, längst bekannt ist, bietet das SrfAC-Strukturmodell den ersten Einblick in die Anordnung der einzelnen Domänen, ihre Interaktionen, sowie die verbindenden Linkerbereiche. Dabei bilden die C- und die A<sub>core</sub>-Domäne eine Plattform (*Workbench*), auf der die A<sub>sub</sub>- sowie die PCP-Domäne, während der einzelnen Schritte der Elongation, agieren.

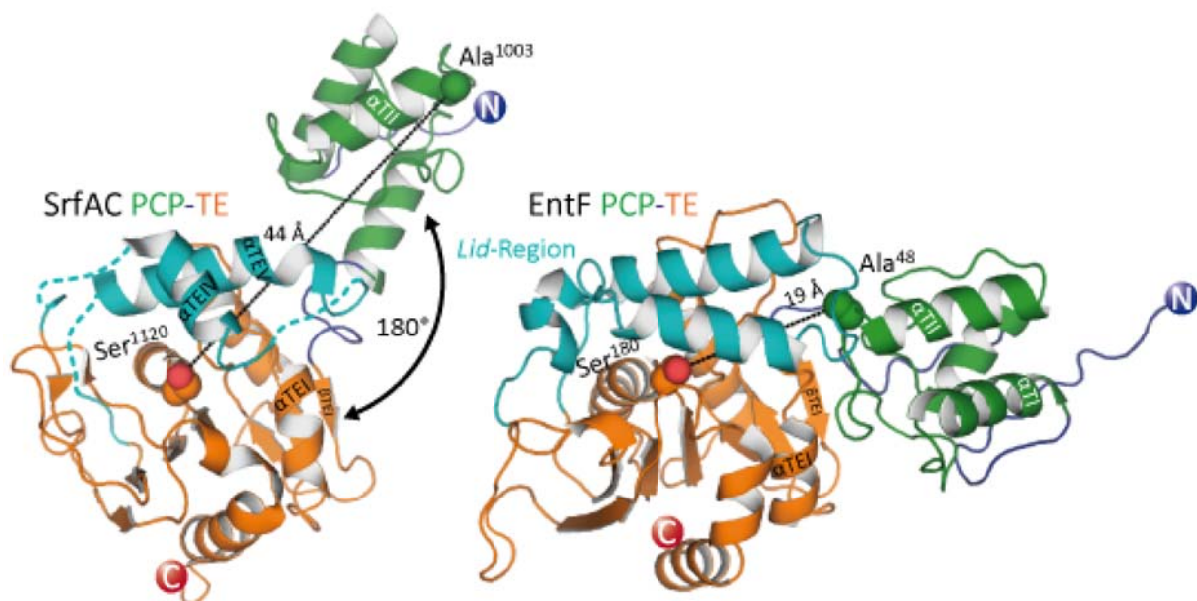


Abbildung 52: SrfAC-EntF Bidomänen-Vergleich durch Superposition der TE-Domäne. Beide PCP-Domänen, die SrfAC-PCP (pdb:2VSQ), sowie die EntF-PCP (pdb:2ROQ), tragen eine Serin zu Alanin Variation des konservierten Serin-Restes, das am Anfang der zweiten  $\alpha$ -Helix lokalisiert ist und zeigen die AH-Konformation der PCP-Domäne. Die EntF PCP-TE Bidomäne zeigt die produktive Interaktion zwischen der TE- und der PCP-Domäne, durch den Loop-Bereich der PCP-Domäne zwischen der  $\alpha$ TI- und  $\alpha$ TII-Helix und der  $\alpha$ TEI-Helix, sowie zusätzliche flankierende Interaktion durch  $\beta$ TEI und dem  $\alpha$ TEIV- $\alpha$ TEV-Loop-Bereich (*Lid-Region*). Die Distanz zwischen Alanin<sup>48</sup> und Serin<sup>180</sup> liegt mit den 19 Å in der Reichweite des PPant-Armes. Die beiden Domänen in dem SrfAC Strukturmodell zeigen keine gemeinsame Interaktionsfläche, die Distanz zwischen Alanin<sup>1003</sup> und Serin<sup>1120</sup> beträgt 44 Å. Die PCP-Domäne ist dabei von der TE-Domäne abgewandt, so dass die  $\alpha$ PII-Helix im Vergleich zu der EntF-PCP um fast 180° gedreht ist. Auch die  $\alpha$ TEIV-Helix der SrfAC-TE ist im Vergleich zu EntF-TE um ca. 45° entgegen dem PPant-Kanals zwischen Alanin<sup>48</sup> und Serin<sup>180</sup> (EntF) gedreht.

### 6.2.1 PCP-Domäne: Arretierung durch rationale Mutagenese

Sobald die Seitenkette des konservierten Serins der PCP-Domäne mit dem PPant-Arm kovalent verbunden ist, kann diese zwischen der AH- bzw. H-Konformation wechseln. Da bislang alle strukturellen Untersuchungen die PCP-Domäne in der AH-Konformation zeigten, wurde durch die Serin zu Alanin Mutante die PCP-Domäne in dieser Konformation

arretiert[111]. Nur bei der SrfAC-S1003A-Variante konnte Kristallbildung beobachtet werden, die PCP-Domäne wurde in der AH-Konformation vorgefunden. Obwohl in dem Zustand eine Interaktion mit der C-Domäne festgestellt wurde, kann die PCP-Domäne durch die flankierenden Linkerbereiche in die Reichweite aller aktiver Zentren gelangen. Dabei beträgt die Distanz des Alanin<sup>1003</sup> zu dem aktivem Zentrum der TE-Domäne (Serin<sup>1120</sup>) 44 Å und zu dem in der A-Domäne gebundenen Leucin 57 Å. Durch die Interaktion der PCP-Domäne mit dem korrespondierenden Partner, der C- (SrfAC) bzw. TE-Domäne (EntF)[273], können in der Reichweite des PPant-Armes (max. 20 Å) die möglichen Interaktionspartner identifiziert werden.

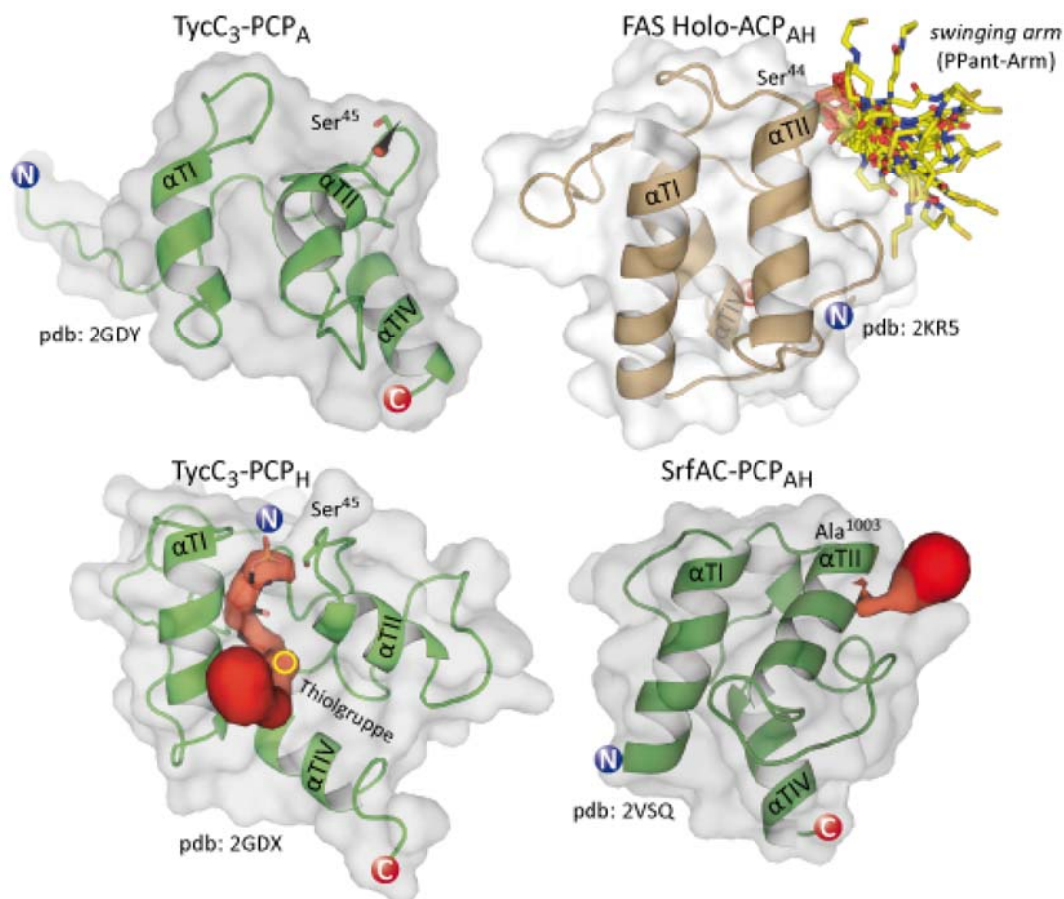


Abbildung 53: Konformationen der TycC<sub>3</sub>-PCP, der SrfAC-PCP und der FAS Holo-PCP. Die apo-PCP kann die A- (pdb: 2GDY) und AH-Konformation (pdb: 2GDW) und die holo-PCP die AH- und die H-Konformation (pdb: 2GDY) annehmen. Die Alanin<sup>1003</sup> Mutante aus SrfAC ist in der AH-Konformation arretiert. Durch die rot gefärbten Kavitäten ist der von Serin<sup>45</sup> bzw. Alanin<sup>1003</sup> zugängliche Raum dargestellt. Die Koordinaten des PPant-Armes in der TycC<sub>3</sub>-PCP<sub>H</sub> Kavität wurden aus der Struktur der Phosphopantothein-Adenyltransferase (pdb: 3K9W) übernommen. Die Holo-ACP aus der Norsolorinsäure-Synthase (NSAS; pdb: 2KR5) zeigt die möglichen Orientierungen des am Serin<sup>44</sup> kovalent gebundenen PPant-Arms. Der PPant-Arm kann so in der AH-Konformation das aktive Zentrum anderer Domänen erreichen, während die Thioester-Bindung in der H-Konformation innerhalb der PCP-Domäne liegt und somit vor Hydrolyse geschützt ist.

Für eine katalysierte Reaktion kann das PPant-gebundenen Substrat eine Reichweite von 16 Å (C-Domäne; SrfAC) bzw. 19 Å (TE-Domäne; EntF) überbrücken. Dafür muss der PPant-Arm exponiert werden, um so das aktive Zentrum erreichen zu können (Histidin<sup>144</sup> in SrfAC bzw. Serin<sup>180</sup> in EntF). Der signifikanteste Unterschied zwischen den zwei möglichen Holo-PCP Konformationen ist die Orientierung und Länge der  $\alpha$ TII-Helix. In der AH-Konformation wird sie durch interne Wechselwirkungen des konservierten Serins bzw. Alanins, stabilisiert und verläuft parallel zu der  $\alpha$ TI-Helix[227]. Die  $\alpha$ TII-Helix in der H-Konformation dagegen ist beinahe orthogonal zu der  $\alpha$ TI-Helix, das konservierte Serin bzw. Alanin befindet sich exponiert in einem Loop unmittelbar davor. Dadurch wird die Position und die Orientierung des Serins und somit auch des PPant-Armes stark beeinflusst. Mithilfe der Software Carver 2.1.2[274] können die verfügbaren Hohlräume ausgehend von einem Startpunkt (Atom, Heteroatom, Aminosäure oder ganze Selektionen) berechnet und in Form einer Kavität dargestellt werden. Die berechneten Kavitäten für die unterschiedlichen Konformationen der PCP-Domäne sind in der Abbildung 53 dargestellt, ausgehend vom Serin bzw. Alanin als Startpunkt. In der H-Konformation befindet sich zwischen der  $\alpha$ TI- und  $\alpha$ TII-Helix eine deutliche Kavität. Bei einem in dieser Kavität positionierte, PPant-Arm (Koordinaten aus pdb: 3K9W übernommen) würde die Thiolgruppe vor Hydrolyse geschützt im Inneren der PCP-Domäne liegen. Der PPant-Arm, gebunden an die PCP-Domäne in der AH-Konformation, wäre, wie bereits für die FAS-ACP gezeigt[275], frei zugänglich (Abb. 53). Daher kann die PCP-Domäne für eine produktive Interaktion mit einer anderen NRPS-Domäne, deren Interaktionsfläche im Wirkungsradius des PPant-Armes liegt (<20 Å), nur die AH-Konformation annehmen.

### 6.2.2 Orientierung der A<sub>sub</sub>-Domäne als Indikator der Adenylierungsreaktion

Mitglieder der ANL-Familie[97], die katalytische und strukturelle Homologien teilen, werden in drei unterschiedliche Konformationen klassifiziert, wobei die C-terminale A<sub>sub</sub>-Domäne verschiedene Orientierung in Bezug auf die A<sub>core</sub>-Domäne annimmt (Abb. 54A). Die verschiedenen Core-Motive orientieren sich dabei neu, ohne dass die Faltung der A<sub>sub</sub>-Domäne verändert wird (Abb. 54; grauer Kasten). Auch wenn die Orientierung der SrfAC A<sub>sub</sub>-Domäne im Vergleich zu der PheA-Domäne abweicht, zeigt die Superposition einen



RMSD-Wert von nur 1.06 Å für 87 C<sub>α</sub>-Atome. Die Orientierung der SrfAC-A<sub>sub</sub>-Domäne entspricht keiner der drei definierten Konformationen. Vielmehr nimmt sie eine intermediäre Konformation, die zwischen der offenen und der Adenylierungs-Konformation liegt, an.

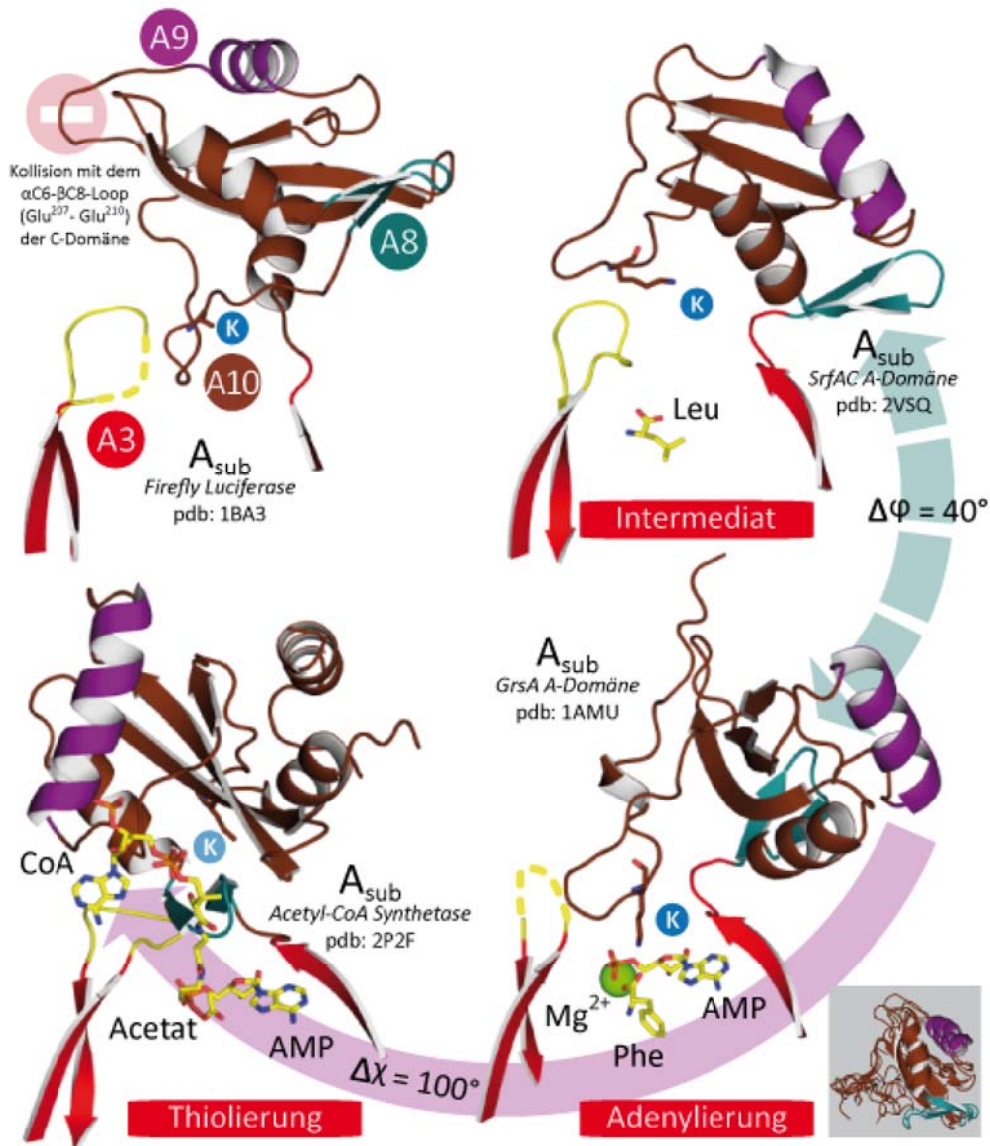


Abbildung 54: Superposition der A<sub>sub</sub>-Domänen aus *Luciferase*, *SrfAC*, *GrsA* und *ACS*. A<sub>sub</sub> Darstellung in der offenen (pdb: 1BA3), Intermediat- (pdb: 2VSQ), Adenylierungs- (pdb: 1AMU) und Thiolierungs- (pdb: 2P2F) Konformation. Die α-Helix (violett; A9-Core Motiv) und das β-Faltblatt (cyan; A8-Core Motiv) vollziehen eine Bewegung von  $\Delta\Phi = 40^\circ$  bzw.  $\Delta\chi = 100^\circ$  zwischen den einzelnen Stationen. Zusammen bilden sie in der Thiolierungskonformation den Pantotheintunnel aus, die Interaktionsfläche für den PPant-Arm sowie die PCP-Domäne. Das Lys948 (braun; A10-Core Motiv) kommt nur bei der Adenylierungskonformation in den Wirkungsradius des gebundenen ATP. Die β,γ-Position des ATP-Phosphatrestes wird in der Adenylierungskonformation durch den P-Loop komplexiert (gelb; A3-Core Motiv). Die Überlagerung der A<sub>sub</sub>-Domänen (grauer Kasten), zeigt die übereinstimmende Faltung, die jedoch in Bezug zur A<sub>core</sub>-Domäne verschiedene Orientierungen annimmt. In einer Workbench zeigt die A<sub>sub</sub>-Domäne in der offenen Konformation eine Kollision mit dem αC6-βC8-Loop (Glu<sup>207</sup>-Glu<sup>210</sup>), somit kann diese Konformation nur von alleinstehenden A-Domänen angenommen werden, die SrfAC A<sub>sub</sub>-Domäne zeigt eine kompatible Intermediat-Konformation.

Die Superposition der SrfAC- $A_{\text{sub}}$ -Domäne gemäß der Stationen der Adenylierung zeigt, dass die offene Konformation nur in alleinstehenden ANL-Domänen eingenommen werden kann, da sie in der Workbench eine Kollision mit dem  $\alpha\text{C6-}\beta\text{C8}$ -Loop aufweist. Somit definiert die Orientierung der SrfAC- $A_{\text{sub}}$ -Domäne die Intermediat-Konformation für die in einer *Workbench* integrierten A-Domänen. In der durch die *Firefly Luciferase* (pdb: 1BA3) definierten offenen Form bzw. der durch SrfAC definierten Intermediat-Konformation (pdb: 2VSQ), erfolgt die Substratkomplexierung (Aminosäure, ATP und  $\text{Mg}^{2+}$ ). Anschließend findet die reversible Adenylierungsreaktion am komplexierten Substrat statt, bei der sich die  $A_{\text{sub}}$ -Domäne um ca.  $40^\circ$  ( $\Delta\phi$ ) zur  $A_{\text{core}}$ -Ebene neigt (Abb. 54), definiert durch die Gramicidin-A-Domäne (PheA, pdb: 1AMU). Während der Adenylierungsreaktion wird die  $\alpha$ - sowie die  $\beta,\gamma$ -Position des ATP-Phosphatrestes vom Lysin (Core-Motiv A10) bzw. dem P-Loop (Core-Motiv A3) komplexiert. Im letzten Schritt, der Thiolierung, wird das Substrat als Thioester am PPant-Arm gebunden. Hierfür vollführt die  $A_{\text{sub}}$ -Domäne eine Bewegung um ca.  $100^\circ$  ( $\Delta\chi$ ) entlang der  $A_{\text{core}}$ -Ebene (Abb. 54). Für diesen, letzten Schritt des Zyklus ist die Interaktion mit dem PPant-Arm sowie der PCP-Domäne notwendig. Dieser Schritt ist nicht reversibel, die Thioesterbindung kann nur durch Hydrolyse oder Kondensation wieder gespalten werden. Dabei interagiert der  $\beta$ -Loop (Core-Motiv A8) mit dem PPant-Arm, während die  $\alpha$ -Helix (Core-Motiv A9) die Fläche für die PCP-Interaktion bildet[108]. Nach der Substratthiolierung nimmt die  $A_{\text{sub}}$ -Domäne die ursprüngliche Intermediat- bzw. Offene-Konformation an. Die Position des SrfAC-Core-Motivs A8 in der  $A_{\text{sub}}$ -Domäne ist im Vergleich zu PheA (pdb: 1AMU) um  $35^\circ$  verdreht, was zu einer Bewegung des katalytisch aktiven Loops (Core-Motiv A10) um  $16 \text{ \AA}$  aus der aktiven Tasche der  $A_{\text{core}}$ -Domäne führt. Die aktive Tasche in der SrfAC A-Domäne ist durch die korrespondierende Aminosäure Leucin belegt. Die zusätzliche ATP-Bindung könnte die Umoorientierung der  $A_{\text{sub}}$ -Domäne auslösen, vergleichbar mit PheA (Abb. 54A).

### 6.2.3 Kondensationsdomäne als NRPS-Schnittstelle

Die Kondensationsdomäne in dem SrfAC Strukturmodell zeigt vier voneinander unterschiedliche Interaktionsflächen (Abb. 55) und ist somit die zentrale Schnittstelle in einer NRPS-Synthetase. Drei der Interaktionsflächen sind in dem SrfAC-Strukturmodell durch ihre korrespondierenden Partner bzw. artifiziellen Strukturelement (*myc*-Epitop) besetzt. Um die

Art dieser Interaktion besser verstehen zu können, wurden von allen mit der C-Domäne interagierenden Domänen elektrostatische Oberflächen (E-Maps) berechnet. Die elektrostatischen Oberflächen wurden mit der Software APBS[229] nach der Poisson-Boltzmann-Gleichung berechnet. Die erforderlichen PQR-Daten wurden durch PDB2PQR[276-277] erzeugt, wobei die Kraftfeld-Berechnungen nach dem AMBER-Kraftfeld[278] erfolgten. Dabei wurden die Werte für die Ionenkonzentration auf 0.1 M und für den Ionenradius, gemäß Pauling für  $O^{2-}$  auf 140 pm gesetzt. Die E-Maps wurden anschließend in den Grenzen von  $\pm 5 k_b T e_c^{-1}$  durch PyMol dargestellt von negativ (rot, -5) über neutral (weiß, 0) bis positiv (blau, +5). Dabei ist die Boltzmann-Konstante ( $1.3806504 \cdot 10^{-23} \text{ JK}^{-1}$ ) durch  $k_b$ , die Temperatur in K durch  $T$  und die Elektronenladung ( $1.60217646 \cdot 10^{-19} \text{ C}$ ) durch  $e_c$  dargestellt.

Die N-terminale Schnittstelle (Interaktionsfläche) ist durch das artifizielle Pendant der korrespondierenden COM-Helix[133], die *myc*-Helix[142] besetzt. Durch die Beschaffenheit des COM-Hand-Motivs ergeben sich zwei unterschiedlich geladene Bereiche. Der Bereich, der im Zentrum des Motivs die Bodenfläche bildet, ist unpolar und bildet hydrophobe Interaktionen aus, im Gegenzug sind die herausragenden Seiten des COM-Hand-Motivs positiv geladen (blau) und können so mit der negativ geladenen Seite (rot) des *myc*-Epitops interagieren (Abb. 55). Die gegenüberliegende C-terminale Seite der C-Domäne zeigt starke Wechselwirkungen mit der  $A_{\text{core}}$ - sowie der  $A_{\text{sub}}$ -Domäne. Die Wechselwirkung ist dabei auf eine Fläche ausgedehnt ( $1621 \text{ \AA}^2$ ) und kann nicht nur auf die polaren Wechselwirkungen beschränkt werden. So zeigt keine der Interaktionsflächen, weder auf der C-Domänen- noch auf der A-Domänen-Seite eine signifikant polarisierte Fläche (Abb. 55). Vielmehr handelt sich hier um eine für die Domänen spezifische Wechselwirkung zur Ausbildung einer *Workbench*. Dadurch könnte die Änderung der aktivierten Aminosäure durch den Austausch der A-Domäne die Störung dieser Interaktion verursachen, was bereits durch die starke Abnahme der Aktivität für einige NRPS-Systeme gezeigt werden konnte[74].

Die Orientierung der Akzeptor-PCP in dem SrfAC-Strukturmodell und ihre Wechselwirkung mit der C-Domäne sind essentiell für das Verständnis der NRPS-Syntheselogik. Dabei interagiert der L-förmige Bereich, bestehend aus der  $\alpha\text{TII}$ -Helix und dem T1-Motiv (GG(H/D)S(L/I)) tragenden Loop, mit der  $\alpha\text{C1}$ - sowie der orthogonal dazu positionierten,  $\alpha\text{C10}$ -Helix.

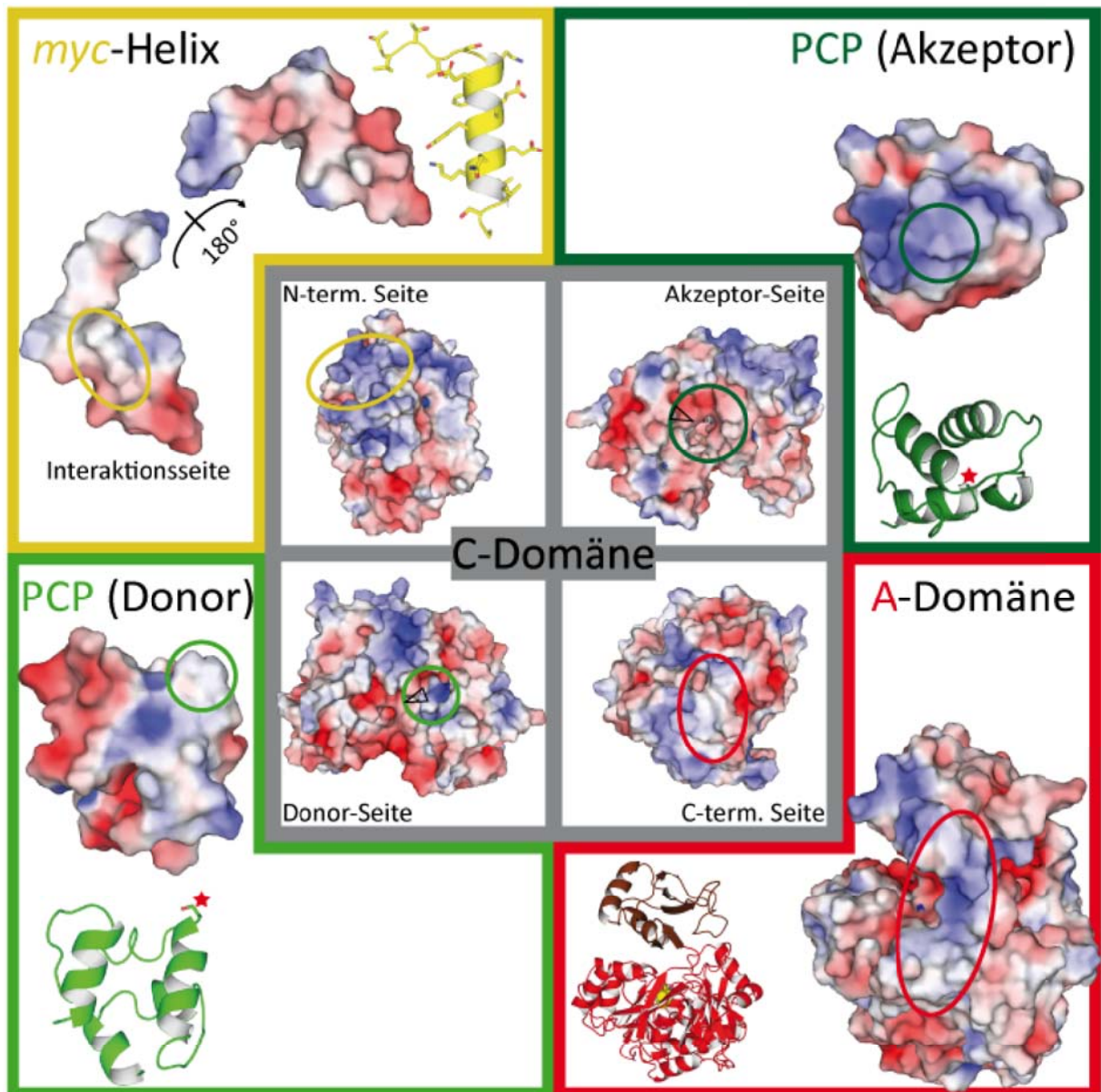


Abbildung 55: Die vier Interaktionsflächen der SrfAC C-Domäne und ihre Interaktionspartner. Die elektrostatische Oberflächen-Darstellung (E-Map) wurde mit der Software APBS nach der Poisson-Boltzmann-Gleichung, die PQR-Daten durch PDB2PQR mit einer 0.1 M Ionenkonzentration und einem Ionenradius von 1.4 Å berechnet. Die E-Maps werden in den Grenzen von  $\pm 5 \text{ k}_B T e_c^{-1}$  dargestellt (von negativ (rot) über neutral (weiß) bis positiv (blau)). Die SrfAC-Kondensationsdomäne und ihre vier Interaktionsflächen im Zentrum der elektrostatischen Oberflächen-Darstellung, sowie die Interaktionspartner in den Ecken. An der N-terminalen Seite ist das COM-Hand-Motiv (gelb umkreist) lokalisiert. Durch die E-Maps werden die unpolare Seitenketten (weiß) sowie die polare an den Seiten (blau) erkenntlich. Die E-Map für die C-terminale Seite zeigt keine spezifischen Flächen, weder an der C-Domänen- noch an der zugewandten A-Domänen-Seite. Die Interaktion mit der PCP-Domäne scheint dagegen durch elektrostatische Wechselwirkungen kontrolliert zu sein. Der schwarze Pfeil markiert den Eingang zum PPant-Eingang. Die Interaktion auf der Donor-Seite wird anhand der modellierten Version der SrfAB-PCP dargestellt, mit der Tyc<sub>5</sub>-PCP als Templat bei 37% Sequenzidentität und einem RMSD-Wert von 2.3. Der hellgrüne Kreis markiert an der PCP-Domäne (durch den roten Stern in der Cartoon-Ansicht markiert) den konservierten Serin-Rest, dessen Umgebung in der PCP-Domäne keine polare Oberfläche zeigt. Neben dem Akzeptor-Eingang der C-Domäne kann eine positiv polarisierte Oberflächenumgebung festgestellt werden. Da die passende negative Ladung in der Umgebung des aktiven Serin-Restes fehlt, konnte sich hierbei um die Interaktionsfläche für den Phosphoryl-Rest des PPant-Armes handeln.



Die Interaktionsflächen beider Domänen zeigen klare elektrostatische Wechselwirkungen, die der PCP-Domäne ist positiv und die der C-Domäne negativ geladen (dunkelgrün umkreist, Abb. 55). Die Interaktionsfläche für die entsprechende Donor-PCP konnte bis jetzt noch nicht identifiziert werden.

### *i.) Oberflächenrepräsentationen und Kavitäten der Kondensationsdomäne*

Bei der Interaktion im TycC<sub>5-6</sub> PCP-C System (pdb: 2JGP) oberhalb der Donor-Öffnung kann eine Distanz zwischen den beiden essentiellen Resten Serin<sup>43</sup> und Histidin<sup>224</sup> von 50 Å gemessen werden, die eindeutig außerhalb der Reichweite des PPant-Armes liegt. Die eindeutige Lokalisierung der Donor-PCP-Interaktionsfläche konnte bisher nicht erfolgen. In der elektrostatischen Oberflächenrepräsentation der SrfAC-C-Domäne ist in dem C-terminalen Bereich unmittelbar neben dem Donor-Zugang eine deutlich positiv geladene Einbuchtung zu sehen (hellgrün markiert; Abb. 55). Das SrfAB-PCP-Domänen-Modell, mit der TycC<sub>5</sub>-PCP (pdb: 2JGP) als Templat bei einer Sequenzidentität von 37% und einem RMSD-Wert von 2.3 für 63 C<sub>α</sub>-Atome, zeigt jedoch im Bereich des konservierten Serin-Restes keine negativ geladene Fläche (Abb. 55). Die Voraussetzung für diese Interaktion ist die kovalente Bindung des PPant-Arms sowie des FA-Peptidyl-Intermediates an der Donor-PCP-Domäne. Die Bindung des PPant-Arms beeinflusst die Oberflächenladung an dem konservierten Serin-Rest durch die kovalente Bindung zu dem Phosphoryl-Rest. Der Phosphoryl-Rest ist essentiell für die Funktion der NRPS-Produktsynthese, daher müsste diese Einbuchtung an der SrfAC-C-Domäne, sollte es sich dabei um die Interaktionsfläche für das Phosphoryl-Rest handeln, auch bei den bereits untersuchten Strukturmodellen der Kondensations- sowie der Epimerisierungsdomänen vorhanden sein.

Es wurden bislang vier NRPS-Domänen strukturell untersucht, die eine gemeinsame Faltung teilen, die TycC<sub>6</sub>- (pdb: 2JGP), VibH- (pdb: 1L5A) und SrfAC- (pdb: 2VSQ) C-Domäne sowie die TycA-E-Domäne (pdb: 2XHG). Alle drei C-Domänen zeigen im C-terminalen Bereich einen Zugang auf der Donor-, sowie der Akzeptor-Seite (Abb. 56). Es zeigt sich eine regelmäßige Oberflächenladung für die positiv geladene Donor-Seite, sowie eine negativ geladene Akzeptor-Seite in allen bisher strukturell untersuchten C-Domänen. Die Oberflächen der Interaktionspartner zeigen die korrespondierende Ladungsverteilung, was auf eine elektrostatische Interaktion hindeutet. Die E-Domäne zeigt allerdings keine Analogien zu der

Ladungsverteilung der C-Domänen. Die Funktionsweise der E-Domäne weicht, bei gleicher Faltung, von der der C-Domänen ab und weist nur einen Interaktionspartner und somit nur die Frontinteraktionsfläche (Donor-Seite) auf.

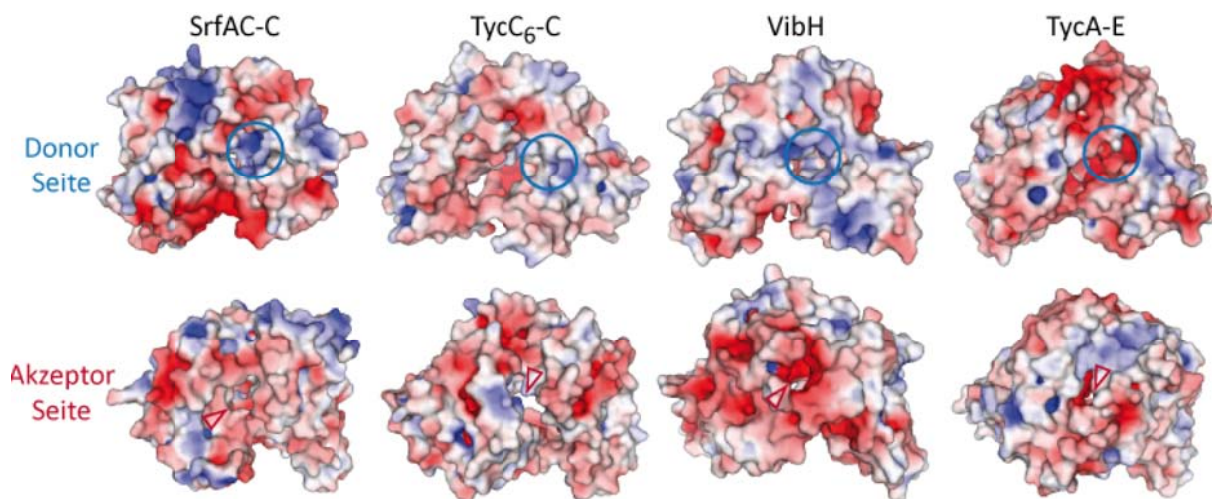


Abbildung 56: Elektrostatische Oberflächenrepräsentation der C- und E-Domänen. Die Darstellung von der Akzeptor- bzw. Donor-Seite für die einzelnen C-Domänen (SrfAC pdb: 2VSQ, TycC<sub>6</sub> pdb: 2JGP und VibH pdb: 1L5A) sowie für die Front- und Rückseite der E-Domäne (pdb: 2XHG). Die Akzeptor-Seite bzw. Rückseite bei der E-Domäne ist durch einen roten Pfeil markiert, die Donor-Seite durch einen blauen Kreis.

Bei genaueren Untersuchungen der unterschiedlichen Domänen zeigen diese zusätzlich verschiedene Kavitäten zwischen den Donor- bzw. Akzeptor-Zugängen und der Reaktionskammer am katalytisch aktiven Histidin-Rest (Abb. 57). Da alle vier Domänen von der Donor-Seite mit dem korrespondierenden Carrier-Protein interagieren kann diese Kavität auch in allen vier untersuchten Strukturmodellen identifiziert werden. Allerdings interagieren nur die SrfAC- und TycC<sub>6</sub>-C-Domänen von der Akzeptor-Seite mit der PCP-Domäne und dem an ihr gebundenen PPant-Arm. Die VibH C-Domäne katalysiert die Kondensationsreaktion zwischen dem PPant-gebundenem 2,3-Dihydrobenzoat (DHB) auf das primäre Amin des freien Norspermidins (NSPD), während die TycA E-Domäne die Epimerisierung des PPant-gebundenen L-Phenylalanins katalysiert und den Akzeptor-Zugang nicht benötigt. Durch die Delokalisation der  $\alpha$ E1-Helix und der *Bridge*-Region im Vergleich zu den C-Domänen scheint sich die grundlegende Form der Kavität innerhalb der E-Domäne von den anderen zu unterscheiden. Die so erfolgte Untersuchung der NRPS-Domänen zeigt die orthogonale Anordnung der einzelnen Zugänge zueinander und die räumliche Ausdehnung der Reaktionskammer am katalytisch aktiven Histidin-Rest.

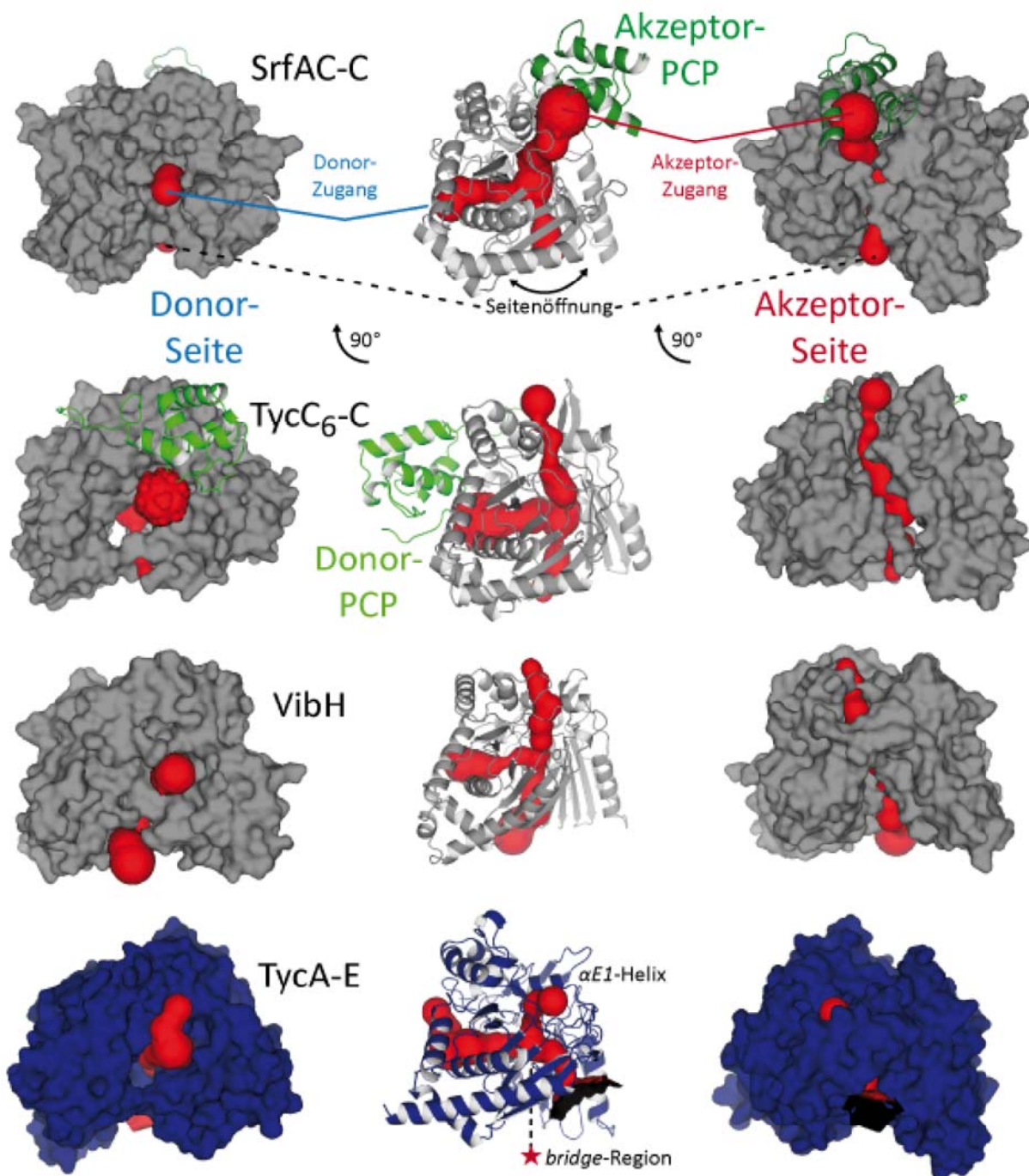


Abbildung 57: Kavitäten der einzelnen Domänen zeigen die Donor- bzw. Akzeptor Öffnungen. Die Ansicht der einzelnen Domänen, SrfAC- (pdb: 2VSQ), TycC<sub>6</sub>-C (pdb: 2JGP) und VibH- (pdb: 1L5A) C-Domäne sowie die TycA-E-Domäne (pdb: 2XHJ), jeweils um 90° rotiert. Die Donor-Ansicht zeigt die Öffnung für den PPant-gebundenen Vorläufer und die Akzeptor-Ansicht die Öffnung für das elongierte Produkt. Während die Donor- und Akzeptor-Kavität in den C-Domänen orthogonal zueinander stehen, zeigt die TycA E-Domäne, durch die Delokalisation der αE1-Helix und der *Bridge*-Region, eine stark deformierten, nicht benötigten Akzeptor-Zugang im Vergleich zu den C-Domänen. Gleichzeitig zeigen die C-Domänen aus SrfAC und TycC<sub>6</sub>, sowie die TycA E-Domäne eine ausgeprägte Reaktionskammer um das katalytisch aktive Histidin, in der die Reaktionen katalysiert wird, sowie eine ausgedehnte Seitenöffnung, die für die Platzierung des Intermediats benutzt werden kann. Die VibH-C-Domäne katalysiert die Kondensationsreaktion zwischen dem PPant-gebundene DHB und freien NSPD und benötigt daher keine räumlich ausgedehnte Kavität für die Vorläuferkette oder den PPant-Arm von der Akzeptor-Seite.

Außerdem zeigt sich ein gravierender Unterschied zwischen dem Donor-Zugang und dem Akzeptor-Zugang, betrachtet man die elektrostatischen Oberflächenrepräsentationen der untersuchten C-Domänen. Während der Donor-Zugang eine positiv geladene Einbuchtung aufweist, um eventuell mit dem Phosphoryl-Rest interagieren zu können, zeigt der Akzeptor-Zugang eine negativ geladene L-förmige Fläche, die bei der SrfAC C-Domäne für die Interaktion mit der korrespondierenden PCP-Domäne verantwortlich ist.

## ii.) *Analogien zwischen den Acetyltransferasen und den NRPS Kondensationsdomänen*

Der Sequenzvergleich der Chloramphenicol und Dihydrolipoyl-Acetyltransferasen mit den NRPS-Kondensationsdomänen kann nicht direkt erfolgen. Der an der Acetyltransferreaktion beteiligte Serin- bzw. Threonin-Rest ist in den CAT-[118] bzw. PDH-E2[279]-Sequenzen stark konserviert. Die NRPS-Kondensationsdomänen weisen keinen solchen konservierten Serin-Rest auf, der im entsprechenden Bereich auf dem strukturell aber nicht sequenziell vergleichbaren, C-terminale C-Subdomäne liegt. Die Superposition der SrfAC-C-Domäne mit den entsprechenden Strukturmodellen zeigt die strukturelle Übereinstimmung des CAT- und E2-Monomers mit dem C-terminalen Bereich der Kondensationsdomäne, dabei fällt der konservierte Serin- bzw. Threonin-Rest der Acetyltransferasen mit dem konservierten Glutaminsäure-Rest aus SrfAC zusammen (Abb. 58). Die in den  $^{\text{D}}\text{C}_\text{L}$ -Kondensationsdomänen hochkonservierte Glutaminsäure ist am Anfang der SrfAC *Bridge*-Region lokalisiert, der flexibelsten Region der Kondensationsdomäne mit den höchsten B-Faktoren. Auch wenn das Glutamin in den  $^{\text{L}}\text{C}_\text{L}$ -Kondensationsdomänen (TycC<sub>6</sub>, pdb: 2JGP) nach dem Core-Motiv C7 kodiert wird, kann in beiden Fällen durch die strukturelle Superposition die entsprechende Aminosäure auf dem vier Aminosäuren langen  $\beta\text{C11}$ -Faltblatt lokalisiert werden. Die strukturelle Ähnlichkeit bricht nach dem  $\beta\text{C11}$ -Faltblatt der *Bridge*-Region ab, da nach diesem Bereich in allen drei Systemen unterschiedliche strukturelle Merkmale folgen.

Analog zu den Phosphopantothein-Transferasen wie ACS aus den FAS-Systemen, welche Transferreaktion als Trimer katalysieren und der PPTase wie Sfp aus NRPS-Systemen die vergleichbare Funktionalität durch ein Pseudodimer erreicht, beruht die strukturelle und funktionelle Verwandtschaft zwischen den Acetyltransferasen und den Kondensationsdomänen der NRPS-Systeme auf einem vergleichbaren Prinzip.



# STRUKTURMODELL DES TERMINATIONSMODULS DER SURFACTIN SYNTHETASE

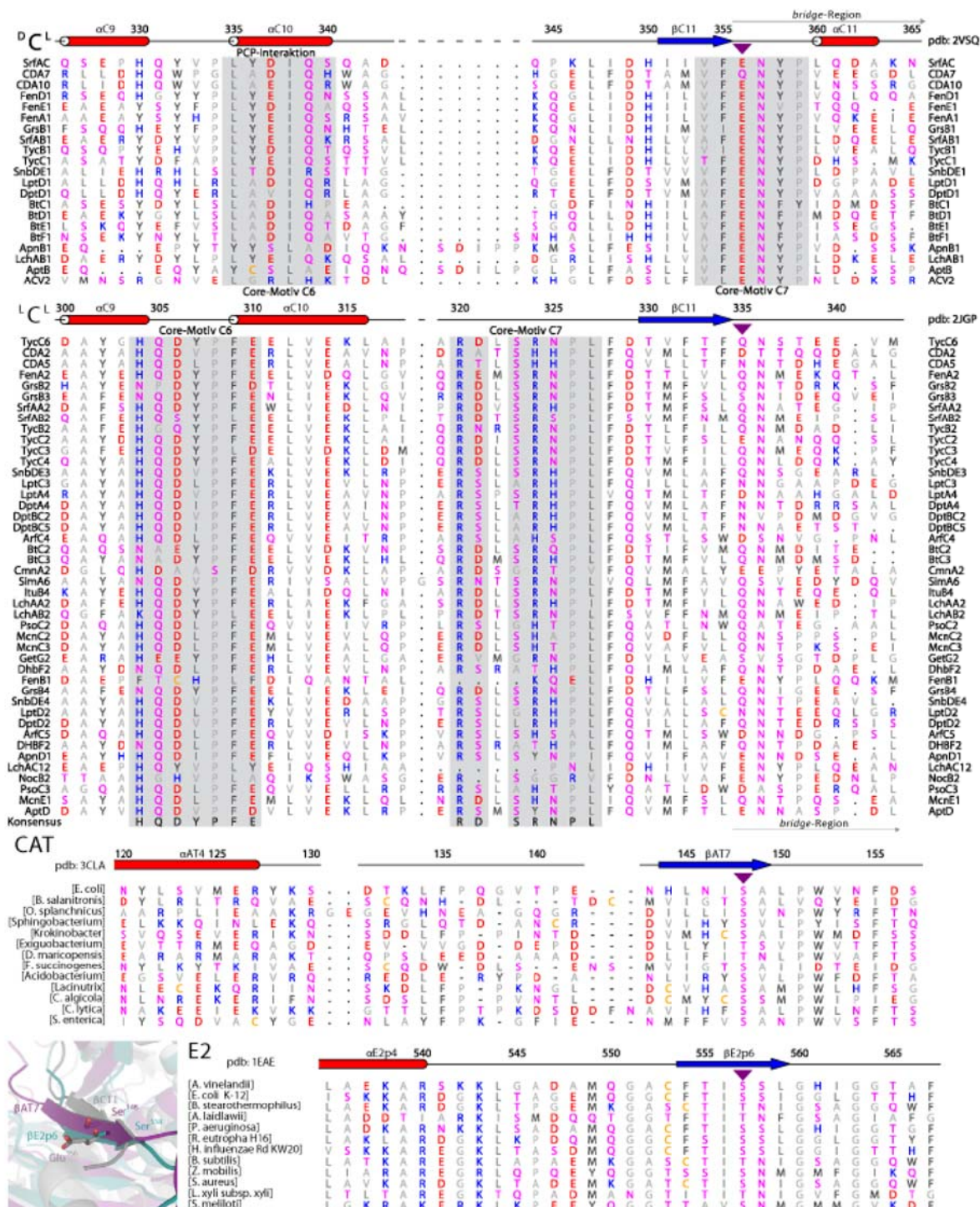


Abbildung 58: Sequenz- und sekundärstrukturvergleich der  $DCL$ -,  $LC$ -, CAT und E2-Domänen. Das Serin bzw. Threonin am 6 Aminosäuren langen  $\beta$ AT7 (CAT) oder  $\beta$ E2p6 (PDH-E2p) ist stark konserviert. Die einzelnen Aminosäuren sind ihrer polaren Eigenschaften entsprechend gefärbt, polar von sauer (rot) bis basisch (blau) bzw. unpolar (grautöne). Der entsprechende Glutaminsäure- bzw. Glutamin-Rest in den  $DCL$ - bzw.  $LC$ -Domänen ist am Anfang der *Bridge*-Region lokalisiert. Dabei zeigt sich zwischen den  $DCL$ - und  $LC$ -Domänen nicht nur ein Unterschied bei dem Glutaminsäure- bzw. Glutamin-Rest, sondern in dem gesamten die Core-Motive C6 und C7 umfassenden Bereich. Die Sequenzen dieser Core-Motive zeigen in ihrer Position und Sequenz eine signifikante Abweichung. Obwohl erst nach dem vier Aminosäuren langen  $\beta$ C11-Faltblatt lokalisiert, zeigt die strukturelle Superposition die räumliche Übereinstimmung zwischen dem Glutaminsäure- bzw. Glutamin-Rest aus NRPS und dem Threonin bzw. Serin in CAT- oder E2-Systemen.

Dabei bilden die Chloramphenicol- und Dihydrolipoyl-Acetyltransferasen als Trimer drei Reaktionskammern aus, die jeweils zwischen den einzelnen Monomeren lokalisiert sind. Diese werden axial von dem Akzeptor-Substrat und äquatorial von den Acetyl-CoA besetzt. Das Substrat und das Acetyl-CoA sind dabei orthogonal zueinander positioniert. In der Reaktionskammer katalysiert der konservierte Histidin-Rest (DHRxx(D/N)G-Motiv) die Deprotonierung des Akzeptor-Substrates, das den Carboxyl-Kohlenstoff am Acetyl-CoA nukleophil angreift. Der resultierende tetraedrische Übergangszustand könnte anschließend von dem Serin- bzw. Threonin-Rest, des am sechs Aminosäuren langen  $\beta$ -Faltblatt des gegenüberliegenden Monomers lokalisiert ist ( $\beta$ AT7 beim CAT und  $\beta$ E2p6 beim PDH-E2p), stabilisiert werden (Abb. 59). Dabei beträgt die Distanz vom Serin-Rest zum Substrat 3.0 Å im E2p- (pdb: 1EAE) bzw. 4.3 Å im CAT-Komplex (pdb: 3CLA). Anschließend wird die Acetylgruppe des komplexierten Acetyl-CoA auf das Substrat übertragen. Sowohl die Chloramphenicol-, als auch die Dihydrolipoyl-Acetyltransferasen wurden bereits mit CoA kokristallisiert und weisen ebenfalls die Bindungsstelle mit der positiven Oberflächenladung für die Bindung der Phosphatgruppe auf (Abb. 59). Auch wenn bei der Superposition von CAT- oder PDH-E2-Komplexen mit dem C-terminalen Bereich der SrfAC-Kondensationsdomäne die konservierte Glutaminsäure<sup>356</sup> und das aktive Serin- bzw. Threonin-Rest räumlich nah positioniert sind, bilden die zwei Bereiche der Kondensationsdomäne einen größeren Winkel, was in einer größeren Entfernung zu dem katalytischen Histidin<sup>147</sup> resultiert. Das erste Strukturmodell einer  $^{\text{D}}\text{C}_L$ -Kondensationsdomäne aus SrfAC ermöglicht detaillierte Einblicke in die NRPS-Kondensationsmechanismen, indem es die vier wichtigen Punkte definiert, die durch die zwei PPant-Arme während der Kondensationsreaktion eingenommen werden können (Abb. 59). Dabei nimmt die Phosphat-Gruppe des PPant-Arm von der Akzeptor-Seite den Ankerpunkt am Alanin<sup>1003</sup> der PCP-Domäne an, während an der Donor-Seite der Ankerpunkt durch die Phosphat-Bindestelle definiert wird. Die jeweiligen PPant-Arme strecken sich dann vom Ankerpunkt bis zum katalytisch aktiven Histidin (His<sup>147</sup>) bzw. der putativ beteiligten Glutaminsäure (Glu<sup>356</sup>). Zusätzlich definieren die berechneten Kavitäten die Geometrie der modellierten PPant-Arme. Die Strukturkoordinaten der PPant-Arme wurden anhand der Form der Kavität aus dem Strukturmodell der Holo-PKS-Malonyl-ACP aus Actinorhodin (pdb: 2KG8) für die Akzeptor-Seite und 4-Hydroxybenzoyl-CoA-



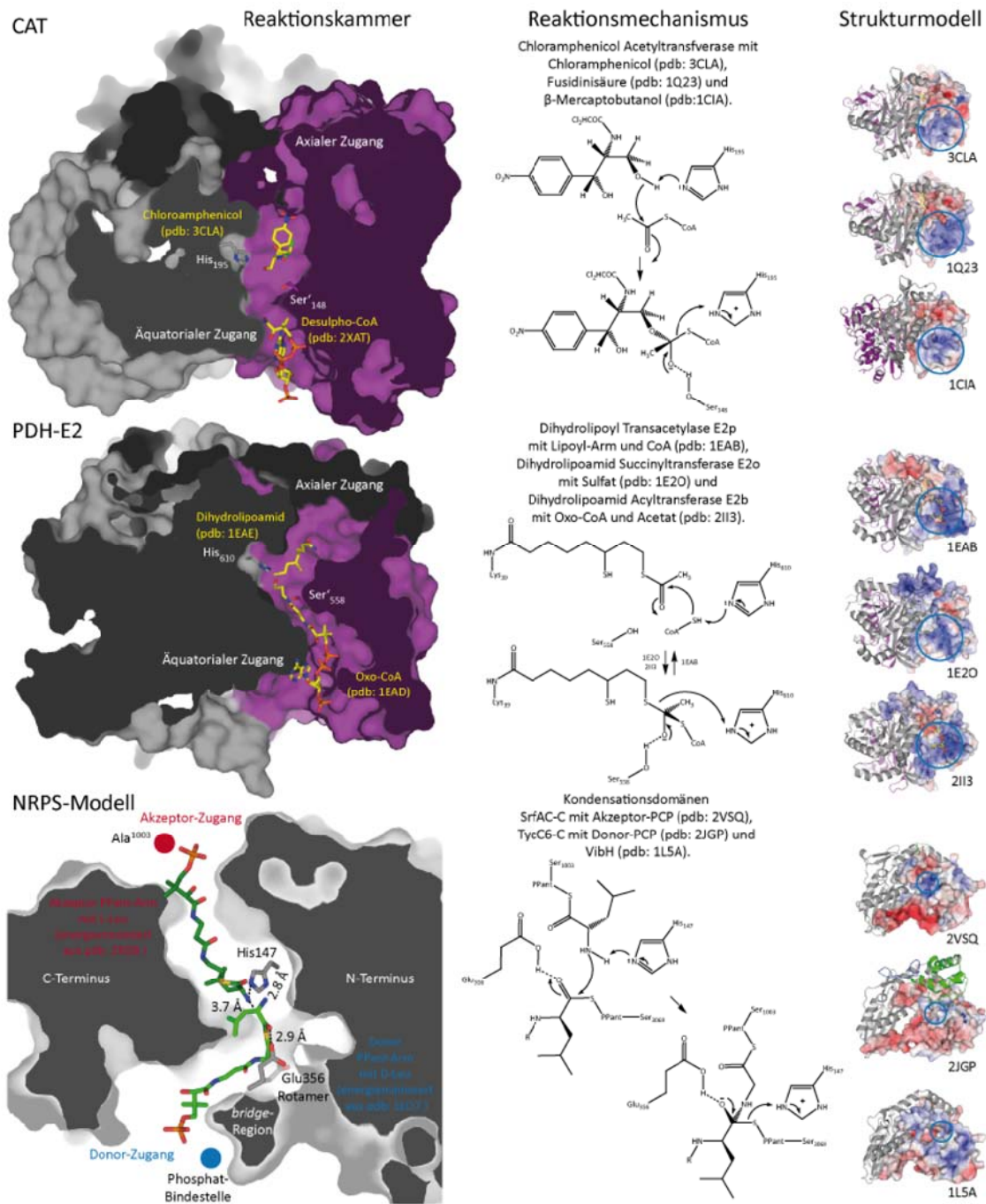


Abbildung 59: Strukturanalogien verwandter Acetyltransferasen mit den NRPS C-Domänen. Die elektrostatischen Oberflächendarstellungen der CAT- bzw. PDH-E2-Komplexe zeigen einen speziellen, positiv geladenen Bereich, für die Bindung von CoA an, der daher einen größeren Radius hat. Durch die besetzten äquatorialen und axialen Zugänge der CAT- und PDH-E2-Komplexe konnten die reaktionsrelevanten Reste identifiziert und der Reaktionsmechanismus postuliert werden. Während in den CAT- und PDH-E2-Komplexen die Reaktionskammer zwischen zwei Monomeren entsteht, wird diese in den NRPS-Kondensationsdomänen als Pseudodimer zwischen dem N- und C-terminalen Bereich gebildet. Die Positionierung der PPant-Arme erfolgt zwischen den vier gekennzeichneten Ankerpunkten Alanin<sup>1003</sup> - Histidin<sup>147</sup> (Akzeptor-Seite), sowie Glutamin<sup>356</sup> und der Phosphat-Bindestelle (Donor-Seite). Die Energieminimierung erfolgte mit UCSF Chimera.



Thioesterase (pdb: 1LO7) für die Donor-Seite ausgewählt. Die so erhaltenen Strukturkoordinaten wurden dann mit den für SrfAC korrespondierenden Aminosäuren (D-Leucin für die Donor- und L-Leucin für die Akzeptor-Seite) verknüpft und anschließend mit dem Programm UCSF Chimera[280] energieminimiert, so dass alle Bindungslängen geometrieoptimiert wurden. Nach der Positionierung befindet sich die L-Leucin-Aminogruppe in der Reichweite des Histidin<sup>147</sup> (2.8 Å), das C<sub>α</sub>-Atom des D-Leucin ist dabei 3.7 Å entfernt. Die Glutaminsäure<sup>356</sup> ist 3.9 Å von dem Carbonyl-Sauerstoff entfernt. Wählt man für die Glutaminsäure<sup>356</sup> das bevorzugte Rotamer, orientiert sich die Carboxylgruppe um und gelangt so in die unmittelbare Nachbarschaft des Carbonyl-Sauerstoffs (2.9 Å). Auch wenn die Distanz zwischen der L-Leucin-Aminogruppe und dem D-Leucin C<sub>α</sub>-Atom 3.9 Å beträgt, würde in dem tetraedrischen Übergangszustand die Distanz verringert werden. In der Elektronendichte der TycC<sub>6</sub>-C-Domäne (pdb: 2JGP) wurde in der Reaktionskammer ein Sulfation bestimmt, das durch Superposition der C-Domänen und der Übertragung in das SrfAC Strukturmodell in der unmittelbaren Nähe der L-Leucin-Aminogruppe (2.0 Å) und dem D-Leucin C<sub>α</sub>-Atom (2.7 Å) positioniert ist. Der Vergleich der katalytisch aktiven Glutaminsäure- bzw. des Glutamin-Rests in den bekannten NRPS-Strukturmodellen, zeigt auf, dass für die zu erfüllende Aufgabe die flexible *Bridge*-Region eine Uorientierung erfahren müsste, damit Glutamin<sup>335</sup> (TycC<sub>6</sub>), Asparagin<sup>335</sup> (VibH) bzw. Glutaminsäure<sup>356</sup> (SrfAC) ihre Funktionalität entfalten können.

Mutationsstudien an der Vibriobactin C-Domäne VibH (pdb: 1L5A) zeigen, dass sich bei der N335L Mutante Umsatzeinbußen bis 80% im Vergleich zum Wildtyp feststellen lassen[117]. Die entsprechende Variation Serin<sup>148A</sup> in dem CAT-System (pdb: 3CLA) führt zu einer 54-fachen Aktivitätsabnahme[118]. Auch wenn diese Variation in der Vibriobactin C-Domäne VibH nicht die vergleichbare Auswirkung im Vergleich zu dem CAT-System aufweist, konnte die katalytische Rolle der Stabilisierung des tetraedrischen Übergangszustandes durch Ser<sup>148</sup>/Ser<sup>558</sup>/Glu<sup>335</sup> bislang weder bestätigt noch widerlegt werden. Mutations- sowie Strukturstudien am PDH-E2-Komplex zeigten, dass die Funktion der Serin<sup>558A</sup> Variation durch andere Aminosäuren-Reste in der Reaktionskammer übernommen werden können[279].

### 6.3 Interaktion in multimodularen Systemen

NRPS-Systeme wie das Cyclosporin-System, die aus nur einer multimodularen Synthetase bestehen, bilden eine Ausnahme. Die Biosynthese vieler NRPS-Systeme erfolgt in mehreren Synthetasen organisiert. Um die Synthesereihenfolge einzuhalten, müssen die entsprechenden Synthetasen durch eine spezifische Interaktion assoziieren, erst dann kann das Produkt assembliert werden. Durch Sequenzuntersuchungen wurde die C-terminale COM-Helix identifiziert, die für die Interaktion von der Donor-Seite aus verantwortlich ist. Diese COM-Helix ist am C-Terminus einer Epimerisierungsdomäne lokalisiert, die um drei Aminosäuren verkürzte Variante zeigte bei Mutationsstudien in der TycA E-Domäne bereits signifikante Aktivitätsverluste, bei längeren Verkürzungen ( $\Delta 6$ ,  $\Delta 9$ ,  $\Delta 12$  sowie  $\Delta 23$ ) konnte keine Aktivität mehr festgestellt werden[133]. In den CAT- und PDH-E2-Komplexen wird durch die Interaktion der Monomere die globale Struktur als Trimer bestimmt. Dabei interagiert das  $\beta AT'1$ -, bzw.  $\beta E'1$ -Faltblatt mit dem  $\beta AT3-4$ -, bzw.  $\beta E3-4$ -Turn und bildet dabei das Drei- $\beta$ -Faltblatt-Motiv aus (Abb. 60). In dem PDH-E2-Komplex interagiert der N-terminale Linkerbereich mit dem Drei- $\beta$ -Faltblatt-Motiv. Das gleiche Motiv, einschließlich der für die Interaktion mit der Akzeptor-PCP verantwortlichen  $\alpha C1$ -Helix, ist in dem N-terminalem Bereich der C-Domäne enthalten. In SrfAC definiert dieses Motiv das COM-Hand-Motiv, das für die Interaktion mit dem bereits beschriebenen COM-Helix-Motiv verantwortlich ist (Abb. 60). Dabei ist die Interaktion zwischen zwei nichtribosomalen Peptidsynthetasen nur dann möglich, wenn diese beiden Motive zueinander kompatibel sind. Sofern andere Auswahlkriterien, z.B. die L-/D-Selektion der C-Domäne, erfüllt sind, kann bei einer Kompatibilität der COM-Motive ein artifizielles Produkt wie beim D-Phe-L-Leu-Dipeptid (TycA-SrfAC-Interaktion) erzeugt werden[155, 281-282]. Die dadurch mögliche Variation würde ein breiteres Spektrum an Sekundärmetaboliten ermöglichen. Ein Beweis für die Effizienz dieser Methode sind die bereits vorhandenen Systeme in lebenden Organismen, die sich dieser Strategie bedienen. Im Cyanobacterium *Anabaena* sp. 90, wurden zwei Start-Synthetasen der Anabaenopeptin Synthetase A (AptA1 und AptA2) identifiziert, die unterschiedliche Vorläufer synthetisieren und aufgrund einer identischen COM-Helix-Sequenz mit der nachfolgenden Synthetase wechselwirken können[241]. Dadurch ist es mit

relativ geringem Aufwand möglich, eine höhere Vielfalt in der NRPS-Produktpalette zu erzeugen[281].

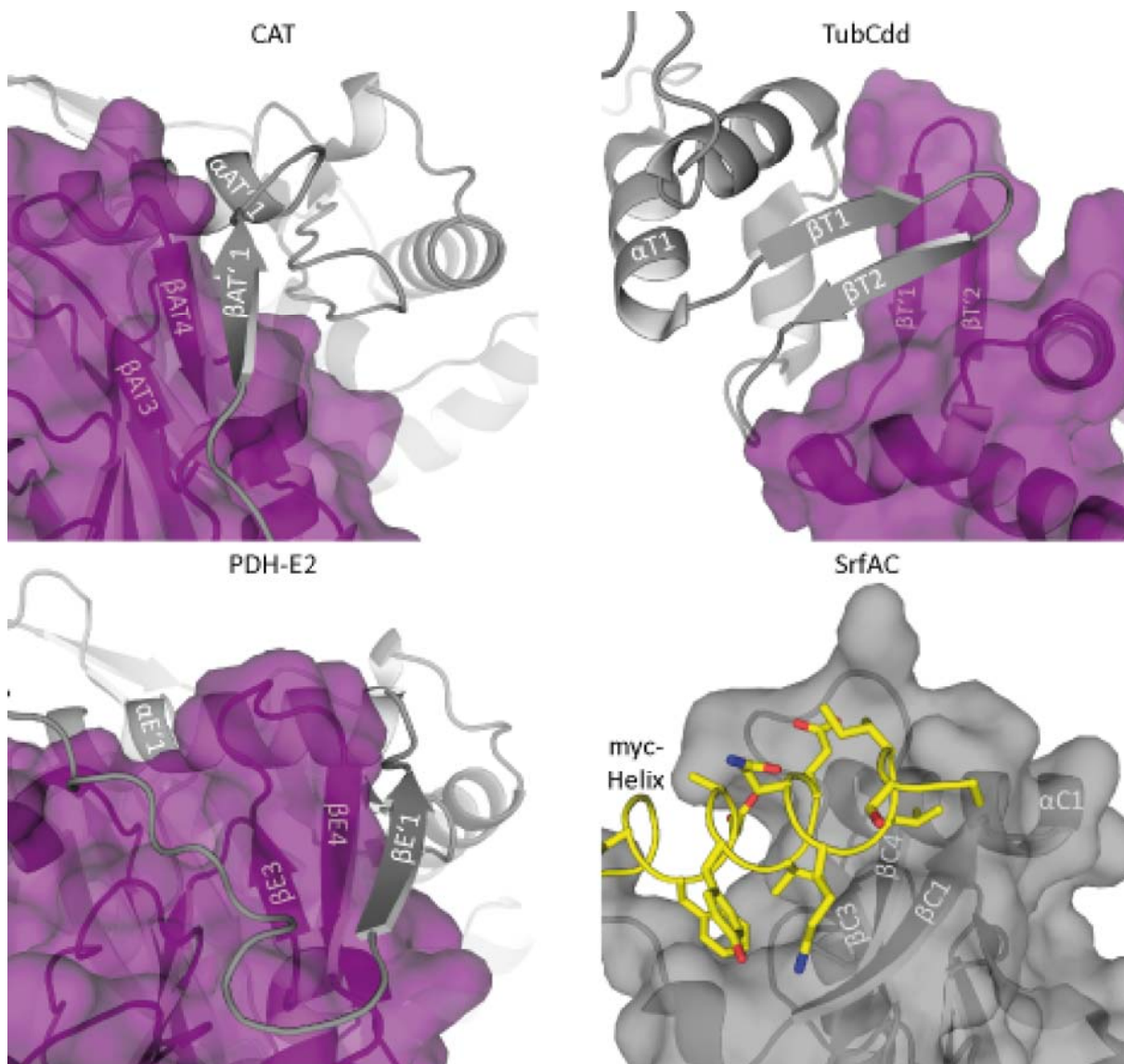


Abbildung 60: Die Interaktionsmotive der CAT-, E2-, PKS-NRPS-Hybrid- und NRPS-Systeme. Bei dem CAT-Komplex (pdb: 2II3) wird das Interaktionsmotiv durch das Drei- $\beta$ -Faltblatt-Motiv ( $\beta$ AT3 und  $\beta$ AT4 und  $\beta$ AT'1) gebildet. Dabei interagiert der  $\beta$ AT3- $\beta$ AT4-Turn mit dem  $\beta$ AT'1-Faltblatt des Nachbarmonomers. Der PDH-E2-Komplex (pdb: 1EAE) zeigt das gleiche Interaktionsmotiv, der N-terminale Bereich interagiert zusätzlich mit dem Drei- $\beta$ -Faltblatt-Motiv. Im Tubulysin PKS-NRPS-Hybridssystem (TubCdd; pdb: 2JUG) wird die Interaktion zwischen den beteiligten Synthetasen durch eine docking-Domäne vermittelt, die in einer einzigartigen  $\alpha\beta\beta\alpha$ -Faltung begründet liegt. Dabei interagiert der  $\beta$ T1- $\beta$ T2-Turn über ein, für den jeweiligen Interaktionspartner spezifisches Motiv. SrfAC (pdb: 2VSQ) zeigt ein anderes, in die C-Domäne eingebettetes und zu den verwandten CAT- und PDH-E2-Komplexen analoges Motiv. Dabei legt sich das COM-Helix-Motiv in das COM-Hand-Motiv, wobei die Interaktion unpolare Kontakte auf dem Boden des Hand-Motives und polare Kontakte auf deren Seiten aufweist.

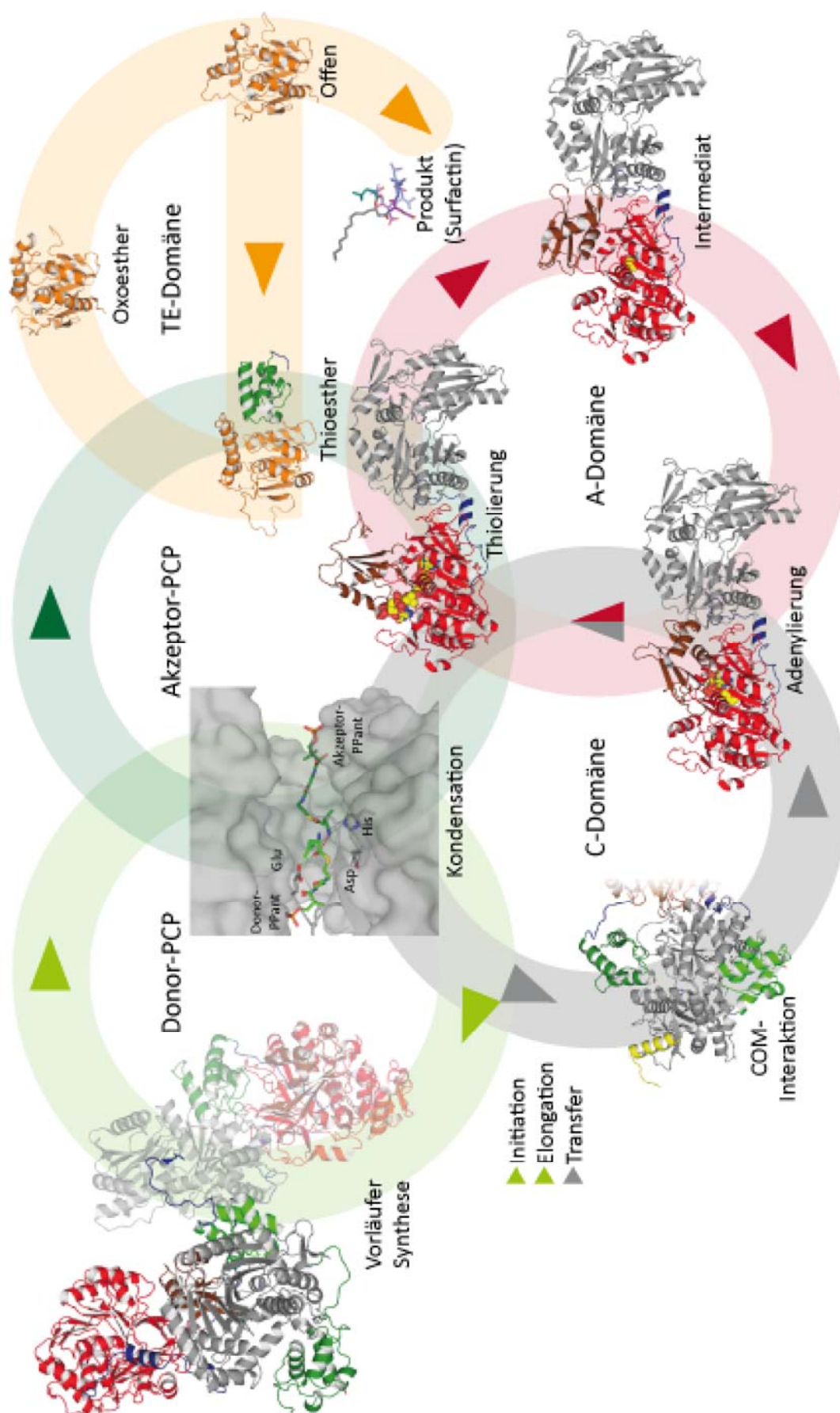
In PKS-NRPS-Hybrid-Systemen zeigt der N-terminale Bereich der Kondensationsdomäne eine geänderte Interaktionsystematik. In dem Tubulysin PKS-NRPS-Hybrid (*Angiococcus*

*disciformis* An d48) wurde eine neuartige Interaktionsystem identifiziert[283]. Das N-terminale COM-Hand-Motiv ist hier um eine Docking-Domäne erweitert. Diese zeigt eine bislang einzigartige  $\alpha\beta\beta\alpha$ -Faltung, wobei die alleinstehende Domäne als Dimer vorliegt (TubCdd pdb: 2JUG). Das Dimer zeigt eine Interaktion über den  $\beta T1$ - $\beta T2$ -Turn, der auch für die Interaktion in dem TubB-TubC-System verantwortlich sein soll (Abb. 60). Dabei soll durch die Ausbildung eines individuellen Interfaces, analog zu der COM-Helix-COM-Hand-Interaktion, die spezifische Interaktion ermöglicht werden.

### 6.4 Überblick: Die Disziplinen der NRPS

Das Surfactin NRPS-System, bestehend aus drei Synthetasen SrfAA (402 kDa), SrfAB (403 kDa) und SrfAC (148 kDa), wurde umfangreich beschrieben und eignet sich deshalb besonders als Modellsystem (Abb. 61). In vorangegangenen Untersuchungen konnten viele Stationen der NRPS katalysierten Synthese, wie unter anderem die Biochemie der Initiationsreaktion[77] oder Struktur der Thioesterasedomäne in der offenen und geschlossenen Konformation[124] geklärt werden.

Abhängig von der Position des Modules in einem NRPS-System ist der erste Schritt entweder die Initiation[284], die Aktivierung und Thiolierung der korrespondierenden Aminosäure durch ein AT-Modul, die Lipoinitiation[77], bei der zuerst eine CoA- oder PCP-gebundene Fettsäure oder der Transfer[133], bei dem der Produktvorläufer von dem vorangegangenen Synthetase über das COM-Motiv auf das CAT-Modul übertragen wird (Abb. 60). Gleichzeitig wird durch die A-Domäne des Moduls die korrespondierende Aminosäure, ATP und  $Mg^{2+}$  (Adenylierungskonformation) gebunden und anschließend auf den PPant-Arm der PCP-Domäne übertragen (Thiolierungskonformation)[108]. Die dynamische PCP-Domäne kann abhängig von dem Beladungs- und Interaktionszustand verschiedene Konformation annehmen[111], dabei kann die Interaktion mit anderen katalytischen NRPS-Domänen nur in der AH-Konformation erfolgen. Die Elongationsreaktion wird nach der Interaktion der Donor- und Akzeptor-PCP mit der Kondensationsdomäne durch diese katalysiert. Durch den Vergleich zu den verwandten CAT- und PDH-E2-Komplexen konnte die Phosphat-Bindestelle identifiziert werden und neben dem Alanin<sup>1003</sup> aus SrfAC die beiden Ankerpunkte für die PPant-Arme bestimmt werden.



Durch die Form der zusätzlich berechneten Kavitäten in der SrfAC Kondensationsdomäne konnten die beiden mit der korrespondierenden Aminosäure beladenen PPant-Arme dargestellt werden. Die L-Leucin-Aminogruppe (Akzeptor-Seite) ist im Wirkungsradius des katalytisch aktiven Histidins (2.7 Å) und gleichzeitig der D-Leucin-Carbonyl-Sauerstoff (Donor-Seite) im Wirkungsradius der Glutaminsäure lokalisiert. In dieser Konstellation befindet sich der D-Leucin-C<sub>α</sub>-Kohlenstoff in räumlicher Nähe zur L-Leucin-Aminogruppe (Abb. 61), so dass die Reaktion stattfinden könnte. Nachdem die Polypeptidkette vollständig assembliert wurde, interagiert die tragende PCP-Domäne mit der Thioesterase und überträgt die als Thioester gebundenes Intermediat auf das katalytisch aktive Serin der Thioesterase[125]. Die *Lid*-Region ist dabei analog zu dem geschlossenen Zustand ausgerichtet, lediglich die *Loop*-Region zwischen den *Lid*-Helices gibt den Zugang zu dem katalytisch aktiven Serin im Inneren frei. Die als Oxoester gebundene Polypeptidkette wird anschließend durch eine Zyklisierungsreaktion oder einer hydrolytischen Abspaltung freigesetzt[122].

Das Prinzip der *Workbench*, das solvatisiert (SAXS) sowie in der Kristallstruktur bestätigt wurde, zeigt die Komplexität in dem Zusammenspiel der NRPS-Domänen, die mehr als nur die lineare Abfolge katalytischer, bzw. dynamischer Domänen sind.

## 7 Ausblick

Durch das unimodulare SrfAC-Strukturmodell konnten wesentliche Merkmale der NRPS-Systeme, wie die *Workbench* und die Linkerbereiche, entschlüsselt werden. Durch das Akzeptor-Modell (*acceptor site only model*) ermöglicht das SrfAC-Strukturmodell im Zusammenspiel mit dem bidomänen TycC<sub>5-6</sub> PCP-C-Strukturmodell einen ersten Einblick in den modularen Aufbau der NRPS-Systeme. Untersuchungen durch Kleinwinkelstreuung (SAXS) zeigen, dass multimodulare NRPS-Gebilde eine wesentlich längere Schraubenachse, im Vergleich zu dem Akzeptor-Modell, besitzen. Zusätzlich zeigen die SAXS-Experimente die Abhängigkeit der globalen NRPS-Struktur vom Beladungszustand der PCP-Domäne. Ohne detaillierte Struktur-Informationen über die einzelnen Konformationen, die durch verschiedene Beladungszustände initiiert werden, konnte die durch die Kleinwinkelstreuung gemessene Elektronendichte nicht zuverlässig durch Strukturmodelle der einzelnen Domänen besetzt werden. Um zuverlässige Modelle für multimodulare NRPS-Systeme zu erhalten, sollten sowohl die Zielproteine als auch die Untersuchungsmethoden der Dynamik der NRPS-Systeme angepasst werden.

### 7.1 Multimodulare NRP-Synthetasen aus thermophilen Organismen

Ein bewährtes Mittel für die strukturellen Untersuchungen an hochdynamischen und heterogenen Komplexen ist die Verwendung der analogen Zielproteine aus thermophilen Organismen[285-286]. Dies wurde bereits für mehrere Systeme, wie FAS aus *Thermomyces lanuginosus*[54] (pdb: 2UVB und 2UVC), Proteasom aus *Thermoplasma acidophilum*[287] (pdb: 3IPM), 70S-Untereinheit des Ribosoms, komplexiert mit Erythromycin aus *Thermus thermophilus*[288] (30S-Bestandteil pdb: 3OHC und 50S-Bestandteil pdb: 3OHD), Substrat-Bindeprotein des ABC-Uptake-Transporters für polare Aminosäuren aus *Streptococcus thermophilus* (pdb: 3HV1; noch nicht publiziert), sowie verschiedene Momentaufnahmen der BV-Monooxygenase aus *Thermobifida fusca*[289] (pdb: 2YLR) gezeigt. Durch diese Vorgehensweise können nicht nur eine bessere Auflösung oder mehrere Konformationen erreicht und verschiedene Substrate bzw. Produkte komplexiert werden, sondern auch die Wahrscheinlichkeit für den Erhalt der Proteinkristalle erhöht werden. Diese kann zusätzlich



durch die Methylierung der Lysinreste im Zielprotein forciert werden[290].

Für die in dieser Arbeit untersuchten trimodularen NRPS-Systeme mit einem Molekulargewicht von über 400 kDa mussten viele Hürden bei der Genexpression, dem Zellaufschluss, der Proteinreinigung und der Beladung überwunden werden. Mit der Reduzierung des Zielproteins auf nur zwei Module, sowie die Wahl des FscG<sub>2-3</sub> aus *Thermobifida fusca* YX (Abb. 62) als Zielprotein, könnten diese Hürden minimiert, die Zugänglichkeit des Zielproteins stark erhöht sowie die Kristallbildung positiv beeinflusst werden. Darüber hinaus könnten verschiedene Beladungszustände der PCP-Domäne, die Serin zu Alanin Variante sowie die Methylierung der Lysinreste die Kristallbildung ebenfalls verbessern.

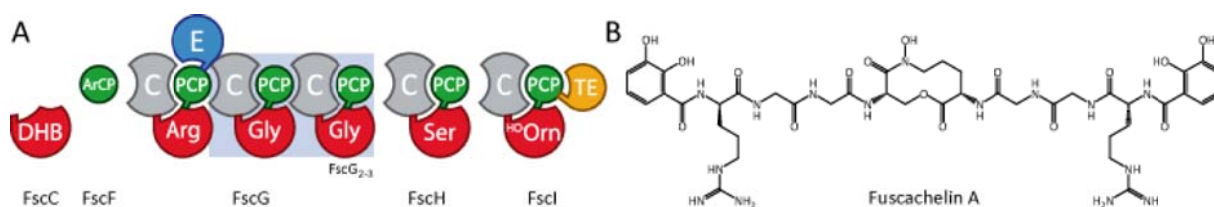


Abbildung 62: Organisation der Fuscachelin Synthetasen sowie die Fuscachelin A-Struktur. (A) Organisation des Fuscachelin-Biosynthese-Clusters aus *Thermobifida fusca* YX. Das bimodulare Zielprotein FscG<sub>2-3</sub> (228 kDa) ist blau umrandet. (B) Produkt der Fuscachelin-Biosynthese: Fuscachelin A.

## 7.2 Cryo-EM an multimodularen NRPS

Der Versuch, mittels der Kleinwinkelstreuung an dimodularen TycB<sub>2-3</sub>-Systemen, generelle Aussagen über die globale NRPS-Struktur zu treffen, scheitert aufgrund der dynamischen Reorganisation der einzelnen Domänen. Auch die Einschränkung der Freiheitsgrade durch den verschiedenen Beladungszustand der PCP-Domänen in den untersuchten NRPS-Systemen ergab nur partiell die benötigte Rigidität. Um diese Freiheitsgrade der modularen Struktur global einzuschränken, wurden NRPS-Systeme mittels der Cryo-Elektronenmikroskopie untersucht. Messbedingt können nur Zielproteine mit einem Molekulargewicht über 300 kDa erfolgreich untersucht werden. Daher wurde das Zielsystem auf das trimodulare SrfAA-Zielprotein (402 kDa) umgestellt.

Die Proben wurden zuerst auf ein beschichtetes und ionisiertes Plättchen (*grid*) aufgetragen und anschließend in flüssigem Ethin schockgefroren. In der ersten Untersuchung am F20-Elektronenmikroskop (*Tecnai*) wurde durch die Kontrastmittel-Färbung (*negative staining*) die Qualität der Proteinprobe überprüft. Anschließend wurde die SrfAA-Probe auf ein als Raster funktionalisiertes Plättchen aufgetragen und erneut am Elektronenmikroskop aufgenommen.

Die aus den Aufnahmen extrahierten Einzelbilder (*images*) wurden mit dem auf SrfAA angepassten NRPS-Akzeptor-Modell als Vorlage auf übereinstimmende Figuren (*figures*) verglichen. Die daraus resultierenden Figuren wurden nach der dem Einzelpartikel entsprechenden Form abgesucht und nach dem Grad ihrer Übereinstimmung und dem Winkel ihrer Projektion klassifiziert (*class*). Dieser Vorgang erfolgte iterativ, wobei im ersten Durchgang ca. 800 – 900 separate Einzelbilder aufgenommen und anhand des vorgegebenen NRPS-Strukturmodells, bei einer Auflösung von ca. 40 Å, nach übereinstimmenden Objekten untersucht wurden. Die Bilder wurden anschließend nach verschiedenen Projektionen des vorgegebenen NRPS-Modell abgesucht und mit diesem superpositioniert (Abb. 63A). Mit Hilfe der Klassen mit der besten Übereinstimmung wurde ein räumliches Modell angefertigt und als neues Modelltemplat benutzt (Abb. 63B). Durch fünf weitere Durchgänge wurden die Projektionen verfeinert und anschließend mit dem Suchmodell verglichen (Abb. 63C).

Der strukturelle Befund zeigt die Übereinstimmung der ermittelten Struktur des Apo-Zielproteins mit dem Strukturmodell, wenn auch eine gewisse Abweichung zu verzeichnen ist. Um diesen strukturellen Befund zu verfeinern, wurde sowohl die Anzahl der aufgenommenen Partikel auf 35000 als auch die der Vergleichsdurchgänge auf 9 erhöht. Die ersten drei Durchgänge wurden diesmal ohne ein Strukturmodell aufgenommen, um ein natives Strukturtemplat zu erhalten (Abb. 64A). Alle aufgenommenen Projektionen wurden aufeinander ausgerichtet und über mehrere iterative Durchgänge gemittelt. Anschließend wurden sie ihrer Projektion entsprechend in Klassen selektiert, sowie unspezifische Klassen für die spätere Betrachtung entfernt (Abb. 64B). Diese Klassen stehen für verschiedene Blickwinkel auf das Zielprotein. Durch die einzelnen Übereinstimmungen zwischen den verschiedenen Klassen (*common lines orientation search*) wurde zuerst die absolute Orientierung der entsprechenden Klasse bestimmt und anschließend das räumliche Modell erstellt. Mit diesem Modell wurde die iterative Suche für weitere sechs Durchgänge durchgeführt (Abb. 64A). Nach der letzten Wiederholung wurde das räumliche Modell anhand des Apo-SrfAA erstellt (Abb. 64C). Das erstellte Modell hat gewisse globale Ähnlichkeiten mit dem NRPS-Modell, wie die schraubenförmige Ausdehnung und die periphere Verteilung von Domänen (Abb. 64C). Jedoch zeigt das erstellte Modell auch Unterschiede zu dem Vergleichsmodell, es werden mehr Domänen detektiert als erwartet.

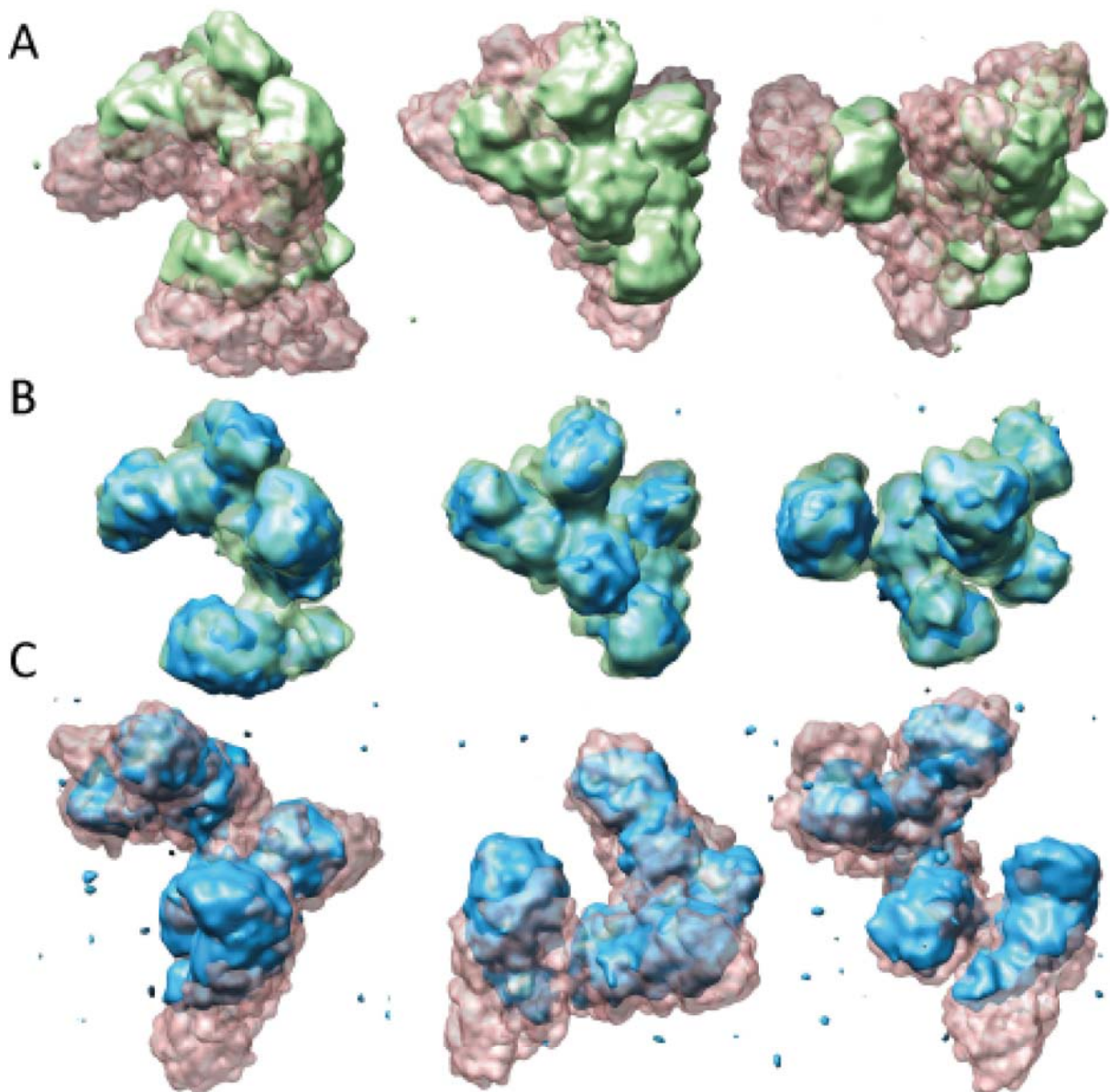


Abbildung 63: Cryo-Elektronenmikroskopie Aufnahmen im Vergleich mit den NRPS-Modell. (A) Überlagerung der Ergebnisse aus dem ersten Suchdurchgang (grün) mit dem NRPS-Modell (rosa). (B) Vergleich der Ergebnisse des ersten (grün) und fünften (blau) Suchdurchgangs. (C) Gegenüberstellung des NRPS-Modells (rosa) und des Ergebnisses des fünften Suchdurchgangs.

Dies kann mehrere Ursachen haben: zunächst wurde die Probe unter Raumtemperatur vorbereitet, eine Vorbereitung bei 4°C würde die Beweglichkeit der Domänen zueinander zumindestens einschränken. Zusätzlich scheinen die Plättchen (*grids*) eine zu dichte Packung der Zielproteine zu beinhalten, die sich gegenseitig zu beeinflussen scheinen.

Wie bereits durch SAXS-Experimente gezeigt, beschreiben die Zielproteine in unterschiedlichen Beladungszuständen eine voneinander abweichende globale Anordnung und

erhöhte Rigidität gegenüber dem Apo-Protein. Daher wurden die mit CoA-SH, CoA-NH-AA und mit Goldnanopartikeln[291-292] (Durchmesser 1.4 nm) verlinkten Cofaktoren (CoA-Au) auf die PCP-Domäne beladen und für die Cryo-Elektronenmikroskopie Analyse vorbereitet.

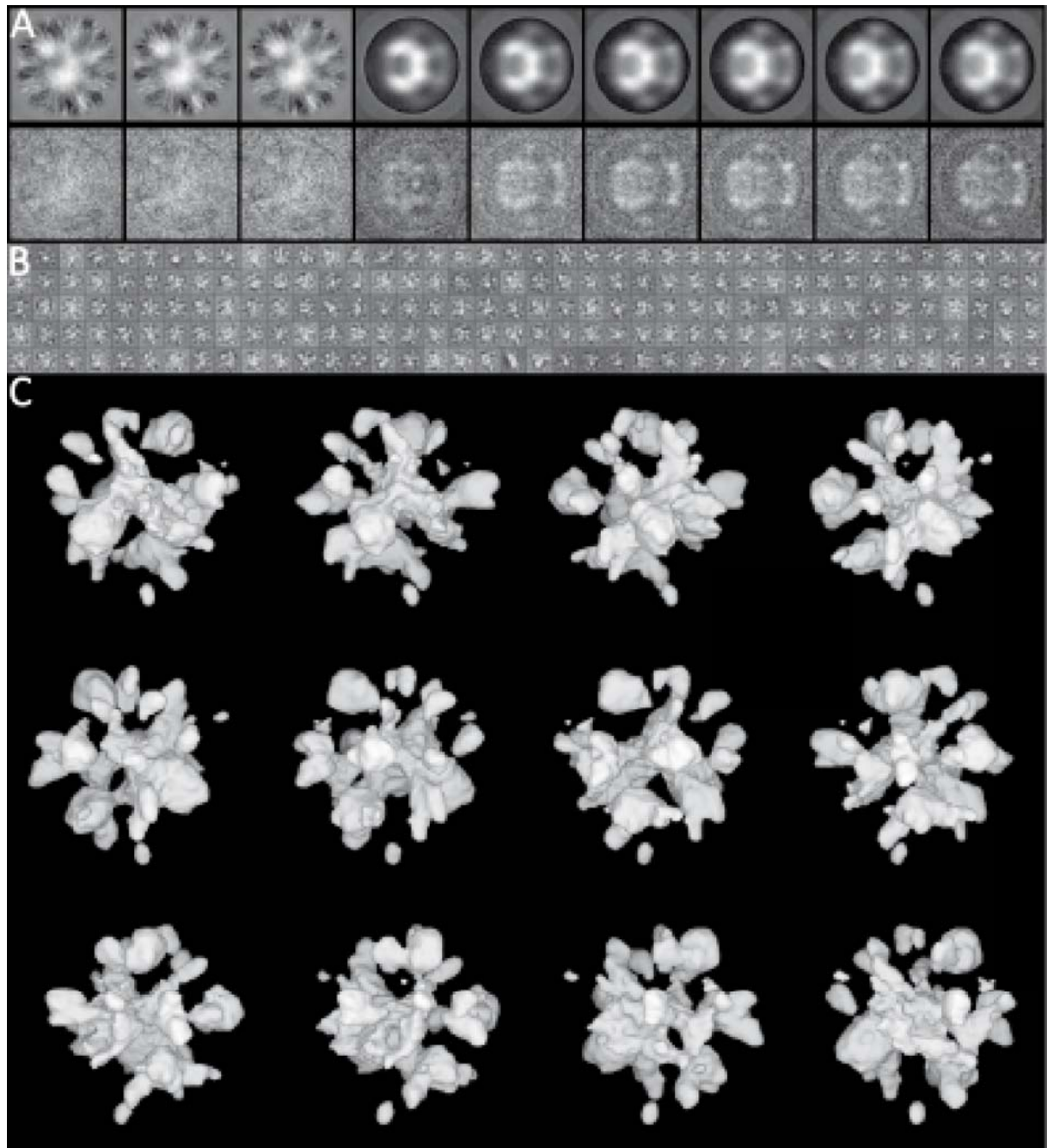


Abbildung 64: Modellfreie, iterative Modellrekonstruktion und deren räumliche Projektionen. (A) Neun iterative Suchdurchgänge, durch die ersten drei wurde das Strukturtemplat für die weitere Suche erstellt. (B) In verschiedene Klassen eingordnete Projektionen für die Modellerstellung. (C) Durch die unterschiedlichen Klassen erstellte räumliche Modelle

Für die Präparation der Amino-CoA-Au-Probe wurde das Amino-CoA nach Herstellerangaben mit der Nanogold-Reagenz inkubiert. Dabei reagiert der Sulfo-N-Hydroxysuccinimid-Ester der Nanogold-Probe mit der freien Aminogruppe des Amino-CoA[293]. Anschließend wurde das Amino-CoA-Au in einem Sfp-Assay auf das SrfAA-Zielprotein übertragen und durch  $\text{Ni}^{2+}$ -NTA von freiem Amino-CoA-Au befreit. Eine ausführliche Analyse aller mit CoA-Derivaten beladenen Zielproteine steht noch an.

## 8 Anhang

### 8.1 Quellcode des Python-Scripts

```
# Simple python program to do basic statistics on fasta files 2010, Alan Tanovic, Ulrich Welling
# usage : python reader.py <file with proteins in fasta format> <file with statistics to assemble>
# for example : python reader.py proteins.txt control.txt
```

```
from numpy import *
import sys
```

```
# Open protein file
file = open(sys.argv[1])
```

```
# Arrays for names and sequence data
names = []
sequences = []
```

```
print "#####"
print "Reading sequences"
```

```
while 1:
    # Read a line
    line = file.readline()
    # Check if file is fully read in
    if not line:
        break
    pass
    # New protein
    if (line[0] == '>'):
        names.append(line)
        sequences.append("")
        print "- Sequence ", len(sequences)
    # protein data
    if (line[0] <>'>'):
        # remove linebreaks and add together sequence data
        sequences[len(sequences)-1] += line.strip("\n")
```

```
print " ", len(sequences), "sequences found"
```

```
# data ist now in names[] and sequences[]
```

```
# Read in Control file
blocks = []
blockmotives = []
file2 = open(sys.argv[2])
```

```
print "#####"
print "Reading analysis blocks"
```

```
while 1:
    # Read a line
    line = file2.readline()
    # Check if file is fully read in
    if not line:
        break
    pass
    # New block
    if (line[0] != '#'):
        a = []
        motives = line.split()
        blockmotives.append(motives)
        pos = sequences[0].find(motives[0])
        pos = pos + len(motives[0])
        a.append(pos)
        pos = sequences[0].find(motives[1])
        a.append(pos)
        blocks.append(a)
```

```
# analysis blocks are now in blocks[] (blocks[][0] - blocks[][1])

print " ", len(blocks), "blocks found"

# Now do the analysis
results = []

print "#####"
print "Starting analysis"
print
print

# Loop over different blocks to analyze
for x in range(0, len(blocks)):
    # Print the status
    print "#####"
    print blockmotives[x][0], blockmotives[x][1], '(', blocks[x][0], '-', blocks[x][1], ')'
    azidcounts = []
    aminoacids = []
    aminoacids_per_sequence = []

    # Loop over the sequences
    for y in range(0, len(sequences)):
        # Loop from start to end of analysis-block
        counter = 0
        aa_seq = []
        for z in range(blocks[x][0], blocks[x][1]):
            if (sequences[y][z] != '-'):
                counter+=1
                aminoacid = sequences[y][z]
                aminoacids.append(aminoacid)
                aa_seq.append(aminoacid)

        azidcounts.append(counter)
        aminoacids_per_sequence.append(aa_seq)
    data = array(azidcounts)
    # standart deviation is the squareroot of the variance of a set
    print data.mean(), '+/-' ,sqrt(data.var())
    print
    results.append(data)

# Build histogram
n, borders = histogram(data, data.max() - data.min() + 1 , (data.min()-0.5, data.max()+0.5), False, None, True )

print "histogram: x ; y"
z = 0
for x in range (data.min(), data.max()+1):
    print x, ",", n[z]
    z = z + 1
print
print

# Look for sequences that are above or below average +/- standart deviation
for y in range(0, len(sequences)):
    # above mean
    if (data[y] > data.mean()+ sqrt(data.var())):
        print "Unusual candidate with much higher count:"
        print names[argmax(data)].strip('>\n')
        print "with a count of ", data[argmax(data)], "aminoacids"
    # below mean
    if (data[y] < data.mean()- sqrt(data.var())):
        print "Unusual candidate with much lower count:"
        print names[argmin(data)].strip('>\n')
        print "Position",argmin(data),"with a count of", data[argmin(data)], "aminoacids"

print "#####"
print
print "Distribution of aminoacids in this block: (in % of the total ammount of aminoacids)"
print " first row is the average, followed by the sequences in the order defined in the protein file "
print
```



```
aminoacids.append("A")
aminoacids.append("R")
aminoacids.append("N")
aminoacids.append("D")
aminoacids.append("C")
aminoacids.append("E")
aminoacids.append("Q")
aminoacids.append("G")
aminoacids.append("H")
aminoacids.append("I")
aminoacids.append("L")
aminoacids.append("K")
aminoacids.append("M")
aminoacids.append("F")
aminoacids.append("P")
aminoacids.append("S")
aminoacids.append("T")
aminoacids.append("W")
aminoacids.append("Y")
aminoacids.append("V")

aminoacids.sort()
c_acid = aminoacids[0]
counter = 1

for x in range(1, len(aminoacids)):
    if (aminoacids[x] != c_acid):
        acidname = c_acid
        acidcount = counter
        c_acid = aminoacids[x]
        counter = 1
        print "0 ; ", acidname, " ; ", (100.0*(acidcount-1))/(1.0 * (len(aminoacids)-20))
    else :
        counter += 1
print "0 ; ", c_acid, " ; ", (100.0*(counter-1))/(1.0 * (len(aminoacids)-20))

for y in range(0, len(sequences)):
    aminoacids_per_sequence[y].append("A")
    aminoacids_per_sequence[y].append("R")
    aminoacids_per_sequence[y].append("N")
    aminoacids_per_sequence[y].append("D")
    aminoacids_per_sequence[y].append("C")
    aminoacids_per_sequence[y].append("E")
    aminoacids_per_sequence[y].append("Q")
    aminoacids_per_sequence[y].append("G")
    aminoacids_per_sequence[y].append("H")
    aminoacids_per_sequence[y].append("I")
    aminoacids_per_sequence[y].append("L")
    aminoacids_per_sequence[y].append("K")
    aminoacids_per_sequence[y].append("M")
    aminoacids_per_sequence[y].append("F")
    aminoacids_per_sequence[y].append("P")
    aminoacids_per_sequence[y].append("S")
    aminoacids_per_sequence[y].append("T")
    aminoacids_per_sequence[y].append("W")
    aminoacids_per_sequence[y].append("Y")
    aminoacids_per_sequence[y].append("V")

    aminoacids_per_sequence[y].sort()
    c_acid = aminoacids_per_sequence[y][0]
    counter = 1

    for x in range(1, len(aminoacids_per_sequence[y])):
        if (aminoacids_per_sequence[y][x] != c_acid):
            acidname = c_acid
            acidcount = counter
            c_acid = aminoacids_per_sequence[y][x]
            counter = 1
            print y+1, " ; ", acidname, " ; ", (100.0*(acidcount-1))/(1.0 * (len(aminoacids_per_sequence[y])-20))
        else :
            counter += 1
    print y+1, " ; ", c_acid, " ; ", (100.0*(counter-1))/(1.0 * (len(aminoacids_per_sequence[y])-20))
```

## 8.2 Eingabedateien

### 8.2.1 Sequenz-Datei (Alignment)

Die Eingabedatei muss in fasta-Format vorliegen. Alle Sequenzen werden untereinander geschrieben. Zuerst wird der *Header* angegeben, der auf das Zeichen '>' folgt und Informationen über die untenstehende Sequenz enthält. Alle *Gaps* werden mit einem '-' Zeichen gekennzeichnet. Beispiel für die SrfAC-Sequenz aus der COM-Modul-Klasse:

```
>SrfAC   SIII - Leu - COM [Bacillus subtilis]
MSQF-----
-----SKDQVQDMYYLSPMQE
GMLFHAIL--NPGQ---SFYLEQITMKVKGSLNIKCLEESMNVIMDRYDVFRTVFIHE--
KVKRPVQVVLKKRQFHIEEIDLTHLTGSEQTAKINEYKEQDKIR-GFDLTRDIPMRAAIF
K-KAEESFEWVWSYHHIILDGWCFGIVVQDLFKVYNALREQKPYSLPPVKPYKDYIKWLE
-KQDKQASLRYWREYLEGFEQTTFAEQR-KK--QKDGYPEKELLFSPSEAEKAFTELA
KSQHTTLSTALQAVWSVLISRYQQSGDLAFGTVVSGRPAEIKGVEHMGVGLFINVPPRRVK
LSEGITFNGLLKRLQEQLQSEPHQYVPLYDIQSQAD---QPKLIDHIIVFENYPLQDAK
N--EE--SSEN-GFDMV-DVHVFEKSNYDLNLMA--S-PGD-EMLIKLAYNENVFDEAFI
LRLKSQLLTAIQQLIQ-NPDQPVSTINLVDDREREFLLTGLNPPA--QAHE-TKPLTYWF
KEAVNANPDAPALTYSGQTLSEYRELDEEANRIARRLQKHGAGKGSVVALYTKRSLELVIG
ILGVLKAGAAAYLPVDPKLPEDRISYMLADSAAACLLTHQEMKEQAAELPY--TGTT-LFI
DDQTRFEE-----QASDP-ATAIDPNDPAYIMYTS GTTGKPKGNITTHANIQGLVK
-HVDYMAFSDQDTFLSVSNYAFDAFTFDYASMLNAA RLIIADEH-TLLDTERLTDLILQ
ENVNV-MFATTALFNLLTDAG---EDWMKGLRCILFGGERASVP-HVRKALRIM-GPGKL
INCYGPTEGTVFATAHVVDLPDS-ISSLPIGKPISNASVYILNEQSQLQPFGAVGELCI
SGMGVSKGYVNRADLTKEKFIENPFK--PGETLYRTGDLARWLPDGTIEYAGRIDDQVKI
RGHRIELEEEIEKQLQEYPGVKDAVVVADRHESGDASINAYLVNRT-----QLSAE
DVKAHLKKQLPAYMVPQTFTFLDELPLTTNGKVNKRLLPKPDQDQ--LAE EWIGPRNEME
ETIAQIWSEVLGRKQIGIHDDFFALGGHALKAMTAASRIKKELGIDLVPKLLFEAPTIA G
ISAYLKN-----G
```

## 8.2.2 Control-Datei

Die Control-Datei ist eine Textdatei und enthält die Marker zwischen denen die Aminosäuren untersucht werden. Die Marker-Sequenz muss exakt mit dem Bereich der ersten Sequenz des Alignments übereinstimmen, enthaltene *Gaps* müssen auch angegeben werden. Jede Zeile muss durch einen Zeilenumbruch getrennt werden (Notepad++). Beispiel für die ersten drei Intercore-Bereiche der COM-Modul-Klasse:

```
# Count aminoacids different from '-' between the following
positions:
MYYLSPMQ RYDVFRTVF
RYDVFRTVF HHIILDG
HHIILDG YKDYIKW
```

## 8.3 Ausgabedatei

Durch die Eingabe des Befehls:

***python reader.py input.fasta control.txt***

wobei *reader.py* für das Python-Script, *input.fasta* für die Eingabedatei in fasta-Format und *control.txt* für die Control-Datei stehen, wird das Script gestartet.

Durch den erweiterten Befehl:

***python reader.py input.fasta control.txt >output.txt***

werden die Ergebnisse in eine Datei geschrieben. Es werden folgende Ergebnisse ausgegeben:

```
#####
Reading sequences
- Sequence 1
...
- Sequence 22
22 sequences found
#####
Reading analysis blocks
17 blocks found
#####
Starting analysis
#####
MYYLSPMQ RYDVFRTVF ( 119 - 166 )
42.3181818182 +/- 0.819443471696
histogram: x ; y
41 ; 0.0454545454545
42 ; 0.727272727273
43 ; 0.136363636364
```

```
#####
➤ Anzahl der Sequenzen (n)

#####
➤ Anzahl der Marker aus der Control-Datei (A-Z)

#####
➤ Ausgabe der berechneten Werte
#####
➤ Intercore-Bereich zwischen Marker A und B (Position im
Alignment)
➤ Mittelwert und Standardabweichung
histogram: x ; y

➤ Verteilung der Längen als Dezimalprozent
```

44 ; 0.0454545454545  
45 ; 0.0454545454545

Unusual candidate with much higher count:  
SnbDE1 SIII - Pro - COM [Streptomyces pristinaespiralis]  
with a count of 45 aminoacids  
Unusual candidate with much lower count:  
BtE1 SIV - Lys - COM [Brevibacillus texasporus] 63079049  
Position 16 with a count of 41 aminoacids  
#####

Distribution of aminoacids in this block: (in % of the total  
ammount of aminoacids)  
first row is the average, followed by the sequences in the  
order defined in the protein file

0 ; A ; 7.8410311493  
0 ; C ; 0.644468313641  
...  
0 ; Y ; 3.43716433942  
1 ; A ; 2.38095238095  
1 ; C ; 2.38095238095

Unusual candidate with much higher count:  
➤ Header der Sequenz mit dem höchsten und niedrigsten  
Wert.

Distribution of aminoacids in this block: (in % of the total  
ammount of aminoacids)  
➤ Zusammensetzung aller Sequenzen im Alignment (0)  
➤ Zusammensetzung der n-ten Sequenz im Alignment (1-n)

## 8.4 Alignment-Analyse

Eine Übersicht, sowie ein Kladogramm der verwendeten Module ist in der Abbildung 51A zu sehen. Alle C/E-Dualdomänen enthaltenden oder nachfolgenden Module, wie für Arthrofactin Synthetase[242] beschrieben, wurden dabei nicht berücksichtigt. In der Tabelle 19 sind alle berechneten Längen, sowie die dazugehörigen Standardabweichungen angegeben. Die untersuchten Sequenzen, die in drei Klassen unterteilt wurden, die COM-, Elongations- und Terminationsmodule, wurden separat voneinander untersucht. Als Marker wurden die Core-Motive der Kondensations- (C1-C7), Adenylierungs- (A1-A10), der PCP- (T1) und der Thioesterase-Domäne (TE1) verwendet. Zusätzlich wurden die Marker um die Domänen- sowie Subdomänenübergänge und weitere Strukturmerkmale ( $\beta$ C13 – Ende der *Bridge*-Region,  $\alpha$ C13 – C-Terminus der C-Domäne oder N- bzw. C-Terminus einer Domäne) ergänzt. Durch die Analyse kann gezeigt werden, dass zwei Bereiche der Kondensationsdomäne und drei Bereiche der Adenylierungsdomäne eine höhere Sequenzvarianz aufweisen. In der C-Domäne sind das die Bereiche um die innere N-terminale Reaktionskammer (C2 bis C3), sowie der Bereich der die *Bridge*-Region beschreibt (C7 bis  $\beta$ C13 bzw.  $\alpha$ C13, das C-terminale  $\beta$ -Faltblatt bzw.  $\alpha$ -Helix). Beide Bereiche flankieren die Reaktionskammer der C-Domäne und sind so an der katalytischen Reaktion der verschiedenen Produkte beteiligt, was auch die höhere Varianz erklären würde. Der Bereich

der Adenylierungsdomäne, der für die Erkennung und Aktivierung der Substrate zuständig ist (A4 - A5), zeigt neben dem für die Aufnahme ganzer zusätzlicher Domänen bekannten Bereich (A8 - A9) ebenfalls eine höhere Abweichung. Zusätzlich zeigt der von der C-Domäne entlegendste, und so in dem schraubenförmigen NRPS-Akzeptor-Modell exponierte Bereich (A2 - A3) ebenfalls eine höhere Varianz. Dieser Bereich könnte als Interaktionsfläche für die *in trans* wirkenden, modifizierenden Enzyme (*tailoring enzymes*) dienen. Trotz der leicht erhöhten Abweichungen in der Länge der einzelnen Bereiche, kann eine gute Übereinstimmung mit den bereits publizierten Werten festgestellt werden[91]. Die in verschiedenen Klassen geordneten Sequenzen zeigen ähnliche Werte für die Adenylierungsdomäne, während die Werte für die Kondensationsdomäne untereinander eine geringe Variabilität aufweisen. Zusätzlich wurden die die PCP flankierenden Linker, sowie die Thioesterase-Domäne untersucht. Eine sehr hohe Varianz ( $\sigma = 17.0$  Aminosäuren) weist der C-terminale Bereich der Thioesterase auf. Dieser Bereich kodiert unter anderem den sehr flexiblen und in seiner Länge sehr variablen Deckel (*Lid-Region*). Bereits untersuchte Domänen mit der charakteristischen Faltung zeigen, dass dieser Bereich aus einer (Fengycin-TE[294]), zwei (Surfactin-TE[124]), drei (Rifamycin-TE[295]) oder vier (Human-FAS-Thioesterase[51])  $\alpha$ -Helices bestehen kann, was die hohe Diskrepanz erklären könnte. Die Bereiche zwischen dem Core-Motiv der PCP- (T1) und der C-Domäne (C1) beinhalten die gesamten Linkerbereiche (19 Aminosäuren in TycC<sub>5</sub>-PCP-C[121]), können jedoch zusätzliche Sequenzbereiche enthalten, so dass für die Linkerbereiche eine gesonderte Auswertung unternommen wurde.

Tabelle 19: Analyse der Intercore-Bereiche für die einzelnen Alignment-Klassen der COM-, Elongation- und Terminationsmodule. Die Distanz zeigt die verwendeten Marker und das Feld Funktion ihre strukturellen sowie katalytischen Merkmale (siehe Kapitel Einleitung).

Die Core-Motive sind wie in SrfAC bzw. TycC5 vorgefunden angegeben. Die Position gibt den Anfang bzw. das Ende des Intercore-Bereiches an. Anschließend sind der über die unterschiedliche Anzahl von Sequenzen ermittelte Mittelwert (Länge,  $\mu$ ) sowie die Standardabweichung ( $\sigma$ ) in Aminosäuren angegeben. Eine erhöhte klassenübergreifende Abweichung, die auf eine höhere sequenzielle Varianz schliessen lässt, ist hervorgehoben.

In einer Normalverteilung sind immer mindestens 68.2% der ermittelten Längen innerhalb des Intervalls ( $\mu \pm \sigma$ ) enthalten.

Varianz schliessen lässt, ist hervorgehoben.

In einer Normalverteilung sind immer mindestens 68.2% der ermittelten Längen innerhalb des Intervalls ( $\mu \pm \sigma$ ) enthalten.

Tabelle 19: Analyse der Intercore-Bereiche für die einzelnen Alignment-Klassen der COM-, Elongation- und Terminationmodule. Die Distanz zeigt die verwendeten Marker und das Feld Funktion ihre strukturellen sowie katalytischen Merkmale (siehe Kapitel Einleitung).

Die Core-Motive sind wie in SrfAC bzw. Tyc5 vorgedefiniert angegeben. Die Position gibt den Anfang bzw. das Ende des Intercore-Bereiches an. Anschließend sind über die unterschiedliche Anzahl von Sequenzen ermittelte Mittelwert (Länge,  $\mu$ ) sowie die Standardabweichung ( $\sigma$ ) in Aminosäuren angegeben. Eine erhöhte klassenübergreifende Abweichung, die auf eine höhere sequenzielle

Varianz schliessen lässt, ist hervorgehoben.

In einer Normalverteilung sind immer mindestens 68.2% der ermittelten Längen innerhalb des Intervalls ( $\mu \pm \sigma$ ) enthalten.



## 9 Abkürzungsverzeichnis

2,3-Dihydrobenzoat.....DHB	Cryo-Elektronenmikroskopie.....Cryo-EM
4'-Phosphopantotheinyl-Arm.....PPant-Arm	Cystein.....Cys (C)
6-Deoxyerythronolid B Synthase.....DEBS	Dalton.....Da
Acetyl-CoA-Carboxylase.....ACC	Decarboxylase.....E1
Acetyltransferase.....AT	Dehydratase.....DH
Acyl-CoA-Synthetase.....Acs	Density modification.....DM
Acyltransferase.....E2	Deutsches Elektronen Synchrotron.....DESY
Adenylisierungsdomäne.....A-Domäne	Dihydroliponyl-Dehydrogenase.....E3
Alanin.....Ala (A)	Dimerisierungsdomäne.....DM
Allosterische Domäne.....AD	Dimethylsulfoxid.....DMSO
Aminosäure (amino acid).....AA	Dynamic light scattering.....DLS
Arginin.....Arg (R)	Electrospray ionisation.....ESI
Aryl Carrier Protein.....ArCP	Enoylreduktase.....ER
Asparagin.....Asn (N)	Epimerisierungsdomäne.....E-Domäne
Asparaginsäure.....Asp (D)	European Molecular Biology Lab.....EMBL
Asymmetric unit.....ASU	FAS aus Pilzen.....fFAS
Biotin Carboxyl Carrier Protein.....BCCP	FAS aus Säugetieren.....mFAS
Biotin Protein Ligase.....BPL	Fettsäuresynthase.....FAS
Biotin-Carboxylase.....BC	Glutamin.....Gln (Q)
Boltzmann-Konstante.....k <sub>B</sub>	Glutaminsäure.....Glu (E)
Calcium-dependent antibiotic.....CDA	Glycin.....Gly (G)
Calvaminic acid synthase.....CAS	Glycin-Decarboxylase-Komplexes.....GCS
Carboxyltransferase.....CT	Histidin.....His (H)
Chloramphenikol-Acetyltransferase.....CAT	Immobil. Metal Affinity Chrom.....IMAC
Clavamate synthase like superfamily.....CSL	Isoleucin.....Ile (I)
Communication-mediating.....COM	Isopropyl-β-D-thiogalactopyranosid.....IPTG
Contact map analysis.....CMA	Isopropyl-β-D-Thiogalaktosid.....IPTG
Counts per minute.....cpm	Ketoacylreduktase.....KR

Kondensationsdomäne.....C-Domäne	Pyrovat-Carboxylase.....PC
Ladung eines Elektrons.....e <sub>c</sub>	Pyruvat-Dehydrogenase.....PDH
Leucin.....Leu (L)	Reduktase.....R-Domäne
Liponyl Domäne.....LD	Retentionszeit.....RT
Lysin.....Lys (K)	Röntgen-Kleinwinkelstreuung.....
Malonyl-Acetyl-Transferase.....MAT	Small angle x-ray scattering.....SAXS
Malonyl/Palmitoyl-Transferase.....MPT	Root mean square deviation.....RMSD
Methionin.....Met (M)	S-Adenosin-Methionin.....SAM
Molecular replacement.....MR	Serin.....Ser (S)
Monooxidase.....Ox-Domäne	Polyacrylamidgelelektrophorese.....PAGE
Multiple cloning site.....MCS	Short-Chain Dehydrogenase.....SDR
Multiple isomorphous replacement.....MIR	Solid phase peptide synthesis.....SPPS
Multiwavelength anomalous diffraction..MAD	Thiamindiphosphat.....ThDP
N-Methyltransferase.....N-Mt	Thioeserase.....TE-Domäne
Natriumdodecylsulfat.....SDS	Threonin.....Thr (T)
Nichtribosomale Peptidsynthetase.....NRPS	transverse relaxation optimized spectroscopy.
Norspermidin.....NSPD	.....TROSY
Peptide carrier protein.....PCP	Tryptophan.....Trp (W)
Periphere Untereinheit Bindedomäne....PSBD	Tyrosin.....Tyr (Y)
Phenylalanin.....Phe (F)	Valin.....Val (V)
Phenylmethylsulfonylfluorid.....PMSF	Verzweigte β-Hydroxy-Carbonsäure.....BCDH
Phosphopantetheinyl-Transferase.....PPTase	Vierlelektronen Reduktase.....R <sup>4e-</sup> -Domäne
PPTase aus B. subtilis.....Sfp	Wildtyp.....WT
Polyketidsynthase.....PKS	Zweielektronen Reduktase.....R <sup>2e-</sup> -Domäne
Polymerasekettenreaktion.....PCR	(2S,3S)-3-Hydroxyasparagin.....hAsn
Prolin.....Pro (P)	β-Hydroxyglutarat.....OGDH
Protein Model Portal.....PMP	β-Ketoacylsynthase.....KS
Pyrophosphat.....PP <sub>i</sub>	

## 10 Literatur

1. Perham, R. N. (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions, *Annu Rev Biochem* 69, 961-1004.
2. Reed, K. E. and Cronan, J. E., Jr. (1993) Lipoic acid metabolism in *Escherichia coli*: sequencing and functional characterization of the lipA and lipB genes, *J Bacteriol* 175, 1325-1336.
3. Lotierzo, M., Tse Sum Bui, B., Florentin, D., Escalettes, F., and Marquet, A. (2005) Biotin synthase mechanism: an overview, *Biochem Soc Trans* 33, 820-823.
4. Leonardi, R., Zhang, Y. M., Rock, C. O., and Jackowski, S. (2005) Coenzyme A: back in action, *Prog Lipid Res* 44, 125-153.
5. Booker, S. J. (2004) Unraveling the pathway of lipoic acid biosynthesis, *Chem Biol* 11, 10-12.
6. Miller, J. R., Busby, R. W., Jordan, S. W., Cheek, J., Henshaw, T. F., Ashley, G. W., Broderick, J. B., Cronan, J. E., Jr., and Marletta, M. A. (2000) *Escherichia coli* LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein, *Biochemistry* 39, 15166-15178.
7. Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A., and Cronan, J. E. (2003) Assembly of the covalent linkage between lipoic acid and its cognate enzymes, *Chem Biol* 10, 1293-1302.
8. Reed, L. J. (1998) From lipoic acid to multi-enzyme complexes, *Protein Sci* 7, 220-224.
9. Perham, R. N. (1991) Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein, *Biochemistry* 30, 8501-8512.
10. Patel, M. S., and Harris, R. A. (1995) Alpha-keto acid dehydrogenase complexes: nutrient control, gene regulation and genetic defects. Overview, *J Nutr* 125, 1744S-1745S.
11. Berg, A., and de Kok, A. (1997) 2-Oxo acid dehydrogenase multienzyme complexes. The central role of the lipoyl domain, *Biol Chem* 378, 617-634.
12. Harris, R. A., and Paxton, R. (1985) Regulation of branched chain alpha-ketoacid dehydrogenase complex by phosphorylation-dephosphorylation. Minisymposium report, *Fed Proc* 44, 305-315.
13. Vijayakrishnan, S., Kelly, S. M., Gilbert, R. J., Callow, P., Bhella, D., Forsyth, T., Lindsay, J. G., and Byron, O. (2010) Solution structure and characterisation of the human pyruvate dehydrogenase complex core assembly, *J Mol Biol* 399, 71-93.
14. Milne, J. L., Wu, X., Borgnia, M. J., Lengyel, J. S., Brooks, B. R., Shi, D., Perham, R. N., and Subramaniam, S. (2006) Molecular structure of a 9-MDa icosahedral pyruvate

dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy, *J Biol Chem* 281, 4364-4370.

15. Douce, R., Bourguignon, J., Neuburger, M., and Rebeille, F. (2001) The glycine decarboxylase system: a fascinating complex, *Trends Plant Sci* 6, 167-176.
16. Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008) Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia, *Proc Jpn Acad Ser B Phys Biol Sci* 84, 246-263.
17. Zempleni, J., Wijeratne, S. S., and Hassan, Y. I. (2009) Biotin, *Biofactors* 35, 36-46.
18. Jitrapakdee, S., and Wallace, J. C. (2003) The biotin enzyme family: conserved structural motifs and domain rearrangements, *Curr Protein Pept Sci* 4, 217-229.
19. Wallace, J. C., Jitrapakdee, S., and Chapman-Smith, A. (1998) Pyruvate carboxylase, *Int J Biochem Cell Biol* 30, 1-5.
20. Schowen, R. L. (2007) Isotopic and other studies on the molecular origins of substrate regulation of some pyruvate decarboxylases: a reconsideration, *Isotopes Environ Health Stud* 43, 1-16.
21. Carey, P. R., Sonnichsen, F. D., and Yee, V. C. (2004) Transcarboxylase: one of nature's early nanomachines, *IUBMB Life* 56, 575-583.
22. Mattevi, A., Obmolova, G., Kalk, K. H., Teplyakov, A., and Hol, W. G. (1993) Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p), *Biochemistry* 32, 3887-3901.
23. Dardel, F., Davis, A. L., Laue, E. D., and Perham, R. N. (1993) Three-dimensional structure of the lipoyl domain from *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex, *J Mol Biol* 229, 1037-1048.
24. Allen, M. D., Broadhurst, R. W., Solomon, R. G., and Perham, R. N. (2005) Interaction of the E2 and E3 components of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. Use of a truncated protein domain in NMR spectroscopy, *FEBS J* 272, 259-268.
25. Frank, R. A., Titman, C. M., Pratap, J. V., Luisi, B. F., and Perham, R. N. (2004) A molecular switch and proton wire synchronize the active sites in thiamine enzymes, *Science* 306, 872-876.
26. Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N., and Hol, W. G. (1996) Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase, *Structure* 4, 277-286.
27. de Kok, A., Hengeveld, A. F., Martin, A., and Westphal, A. H. (1998) The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria, *Biochim Biophys Acta* 1385, 353-366.
28. St Maurice, M., Reinhardt, L., Surinya, K. H., Attwood, P. V., Wallace, J. C., Cleland, W. W., and Rayment, I. (2007) Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme, *Science* 317, 1076-1079.

29. Jitrapakdee, S., St Maurice, M., Rayment, I., Cleland, W. W., Wallace, J. C., and Attwood, P. V. (2008) Structure, mechanism and regulation of pyruvate carboxylase, *Biochem J* 413, 369-387.
30. Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E., and Flugel, R. S. (1997) Post-translational modification of polyketide and nonribosomal peptide synthases, *Curr Opin Chem Biol* 1, 309-315.
31. Mootz, H. D., Finking, R., and Marahiel, M. A. (2001) 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*, *J Biol Chem* 276, 37289-37298.
32. Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G., and Walsh, C. T. (1997) Ability of *Streptomyces* spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase, *Chem Biol* 4, 17-24.
33. Hiltunen, J. K., Chen, Z., Haapalainen, A. M., Wierenga, R. K., and Kastaniotis, A. J. (2010) Mitochondrial fatty acid synthesis--an adopted set of enzymes making a pathway of major importance for the cellular metabolism, *Prog Lipid Res* 49, 27-45.
34. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases, *Chem Biol* 3, 923-936.
35. Nakano, M. M., Corbell, N., Besson, J., and Zuber, P. (1992) Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*, *Mol Gen Genet* 232, 313-321.
36. Reuter, K., Mofid, M. R., Marahiel, M. A., and Ficner, R. (1999) Crystal structure of the surfactin synthetase-activating enzyme *sfp*: a prototype of the 4'-phosphopantetheinyl transferase superfamily, *EMBO J* 18, 6823-6831.
37. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of *Sfp*, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases, *Biochemistry* 37, 1585-1595.
38. Fichtlscherer, F., Wellein, C., Mittag, M., and Schweizer, E. (2000) A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation, *Eur J Biochem* 267, 2666-2671.
39. Chan, D. I., and Vogel, H. J. (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein, *Biochem J* 430, 1-19.
40. Schujman, G. E., and de Mendoza, D. (2008) Regulation of type II fatty acid synthase in Gram-positive bacteria, *Curr Opin Microbiol* 11, 148-152.
41. Hiltunen, J. K., Schonauer, M. S., Autio, K. J., Mittelmeier, T. M., Kastaniotis, A. J., and Dieckmann, C. L. (2009) Mitochondrial fatty acid synthesis type II: more than just fatty acids, *J Biol Chem* 284, 9011-9015.
42. Maier, T., Leibundgut, M., Boehringer, D., and Ban, N. (2010) Structure and function of eukaryotic fatty acid synthases, *Q Rev Biophys* 43, 373-422.
43. Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008) The multienzyme architecture

of eukaryotic fatty acid synthases, *Curr Opin Struct Biol* 18, 714-725.

44. Byers, D. M., and Gong, H. (2007) Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family, *Biochem Cell Biol* 85, 649-662.
45. Misra, A., Sharma, S. K., Surolia, N., and Surolia, A. (2007) Self-acylation properties of type II fatty acid biosynthesis acyl carrier protein, *Chem Biol* 14, 775-783.
46. Misra, A., Surolia, N., and Surolia, A. (2009) Catalysis and mechanism of malonyl transferase activity in type II fatty acid biosynthesis acyl carrier proteins, *Mol Biosyst* 5, 651-659.
47. Ploskon, E., Arthur, C. J., Kanari, A. L., Wattana-amorn, P., Williams, C., Crosby, J., Simpson, T. J., Willis, C. L., and Crump, M. P. (2010) Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis, *Chem Biol* 17, 776-785.
48. Asturias, F. J., Chadick, J. Z., Cheung, I. K., Stark, H., Witkowski, A., Joshi, A. K., and Smith, S. (2005) Structure and molecular organization of mammalian fatty acid synthase, *Nat Struct Mol Biol* 12, 225-232.
49. Maier, T., Leibundgut, M., and Ban, N. (2008) The crystal structure of a mammalian fatty acid synthase, *Science* 321, 1315-1322.
50. Ploskon, E., Arthur, C. J., Evans, S. E., Williams, C., Crosby, J., Simpson, T. J., and Crump, M. P. (2008) A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains, *J Biol Chem* 283, 518-528.
51. Chakravarty, B., Gu, Z., Chirala, S. S., Wakil, S. J., and Quiocho, F. A. (2004) Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain, *Proc Natl Acad Sci U S A* 101, 15567-15572.
52. Brignole, E. J., Smith, S., and Asturias, F. J. (2009) Conformational flexibility of metazoan fatty acid synthase enables catalysis, *Nat Struct Mol Biol* 16, 190-197.
53. Dixon, N., Wong, L. S., Geerlings, T. H., and Micklefield, J. (2007) Cellular targets of natural products, *Nat Prod Rep* 24, 1288-1310.
54. Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolasek, B., and Ban, N. (2007) Structure of fungal fatty acid synthase and implications for iterative substrate shuttling, *Science* 316, 254-261.
55. Lomakin, I. B., Xiong, Y., and Steitz, T. A. (2007) The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together, *Cell* 129, 319-332.
56. Jenke-Kodama, H., and Dittmann, E. (2009) Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges, *Natural Product Reports* 26, 874-883.
57. Kwan, D. H., and Schulz, F. (2011) The stereochemistry of complex polyketide biosynthesis by modular polyketide synthases, *Molecules* 16, 6092-6115.
58. Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E., and Khosla, C. (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B



- synthase, *Proc Natl Acad Sci U S A* 103, 11124-11129.
59. Keatinge-Clay, A. T., and Stroud, R. M. (2006) The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases, *Structure* 14, 737-748.
  60. Keatinge-Clay, A. (2008) Crystal structure of the erythromycin polyketide synthase dehydratase, *J Mol Biol* 384, 941-953.
  61. Kwan, D. H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J. C., Haydock, S. F., and Leadlay, P. F. (2008) Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases, *Chem Biol* 15, 1231-1240.
  62. Meier, J. L., and Burkart, M. D. (2009) The chemical biology of modular biosynthetic enzymes, *Chemical Society Reviews* 38, 2012-2045.
  63. Katz, L. (2009) The DEBS paradigm for type I modular polyketide synthases and beyond, *Methods Enzymol* 459, 113-142.
  64. Campbell, C. D., and Vederas, J. C. (2010) Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes, *Biopolymers* 93, 755-763.
  65. Khosla, C. (2009) Structures and mechanisms of polyketide synthases, *J Org Chem* 74, 6416-6420.
  66. Katsuyama, Y., Miyazono, K. I., Tanokura, M., Ohnishi, Y., and Horinouchi, S. (2010) Structural and biochemical elucidation of mechanism for decarboxylative condensation of beta-keto acid by curcumin synthase, *J Biol Chem*.
  67. Wilkinson, B., Foster, G., Rudd, B. A., Taylor, N. L., Blackaby, A. P., Sidebottom, P. J., Cooper, D. J., Dawson, M. J., Buss, A. D., Gaisser, S., Bohm, I. U., Rowe, C. J., Cortes, J., Leadlay, P. F., and Staunton, J. (2000) Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase, *Chem Biol* 7, 111-117.
  68. Xue, Y., Zhao, L., Liu, H. W., and Sherman, D. H. (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity, *Proc Natl Acad Sci U S A* 95, 12111-12116.
  69. August, P. R., Tang, L., Yoon, Y. J., Ning, S., Muller, R., Yu, T. W., Taylor, M., Hoffmann, D., Kim, C. G., Zhang, X., Hutchinson, C. R., and Floss, H. G. (1998) Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699, *Chem Biol* 5, 69-79.
  70. Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E., and Khosla, C. (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase, *Chem Biol* 14, 931-943.
  71. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) The 1.8 Å crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*, *Structure* 8, 185-195.
  72. Kapur, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide

B synthase, *Proc Natl Acad Sci U S A* 107, 22066-22071.

73. Mootz, H. D., and Marahiel, M. A. (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains, *J Bacteriol* 179, 6843-6850.
74. Grunewald, J., and Marahiel, M. A. (2006) Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides, *Microbiology and Molecular Biology Reviews* 70, 121-+.
75. Stein, T., Vater, J., Kruff, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H. R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates, *J Biol Chem* 271, 15428-15435.
76. Marahiel, M. A., Krause, M., and Skarpeid, H. J. (1985) Cloning of the tyrocidine synthetase 1 gene from *Bacillus brevis* and its expression in *Escherichia coli*, *Mol Gen Genet* 201, 231-236.
77. Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M., and Marahiel, M. A. (2010) Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation, *Chem Biol* 17, 872-880.
78. Gehring, A. M., Mori, I., and Walsh, C. T. (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF, *Biochemistry* 37, 2648-2659.
79. Schoenafinger, G., Schracke, N., Linne, U., and Marahiel, M. A. (2006) Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin, *J Am Chem Soc* 128, 7406-7407.
80. Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P., and Kuchero, G. (2008) NORINE: a database of nonribosomal peptides, *Nucleic Acids Res* 36, D326-331.
81. Caboche, S., Leclere, V., Pupin, M., Kuchero, G., and Jacques, P. (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity, *J Bacteriol* 192, 5143-5150.
82. Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M. E., and Crosa, J. H. (2008) Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in *Vibrio anguillarum*, *Biometals* 21, 635-648.
83. Konz, D., Klens, A., Schorgendorfer, K., and Marahiel, M. A. (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases, *Chem Biol* 4, 927-937.
84. Schracke, N., Linne, U., Mahlert, C., and Marahiel, M. A. (2005) Synthesis of linear gramicidin requires the cooperation of two independent reductases, *Biochemistry* 44, 8507-8513.
85. Stachelhaus, T., and Walsh, C. T. (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase, *Biochemistry* 39, 5775-5787.

86. Weber, G., Schorgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame, *Curr Genet* 26, 120-125.
87. Silakowski, B., Schairer, H. U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Muller, R. (1999) New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1, *J Biol Chem* 274, 37391-37399.
88. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Curr Opin Chem Biol* 5, 525-534.
89. Samel, S. A., Marahiel, M. A., and Essen, L. O. (2008) How to tailor non-ribosomal peptide products--new clues about the structures and mechanisms of modifying enzymes, *Mol Biosyst* 4, 387-393.
90. Stachelhaus, T., and Marahiel, M. A. (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA, *J Biol Chem* 270, 6163-6169.
91. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, *Chem Rev* 97, 2651-2674.
92. Kavanagh, K. L., Jornvall, H., Persson, B., and Oppermann, U. (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes, *Cell Mol Life Sci* 65, 3895-3906.
93. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases, *ChemBiochem* 3, 491-504.
94. Robbel, L., Hoyer, K. M., and Marahiel, M. A. (2009) TioS T-TE--a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides, *FEBS J* 276, 1641-1653.
95. Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining, *Nature Chemical Biology* 1, 265-269.
96. Strieker, M., Nolan, E. M., Walsh, C. T., and Marahiel, M. A. (2009) Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis, *J Am Chem Soc* 131, 13523-13530.
97. Gulick, A. M. (2009) Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-ribosomal Peptide Synthetases, and Firefly Luciferase, *ACS Chemical Biology* 4.
98. McElroy, W. D., DeLuca, M., and Travis, J. (1967) Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid

utilization are compared, *Science* 157, 150-160.

99. Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990) The P-loop--a common motif in ATP- and GTP-binding proteins, *Trends Biochem Sci* 15, 430-434.
100. Chang, K. H., Xiang, H., and Dunaway-Mariano, D. (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase, *Biochemistry* 36, 15650-15659.
101. Eppelmann, K., Stachelhaus, T., and Marahiel, M. A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics, *Biochemistry* 41, 9718-9726.
102. Conti, E., Franks, N. P., and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes, *Structure* 4, 287-298.
103. Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arai, Y., Sugahara, M., Kuramitsu, S., Yokoyama, S., and Miyano, M. (2004) Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer, *J Biol Chem* 279, 31717-31726.
104. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, *EMBO J* 16, 4174-4183.
105. Gocht, M., and Marahiel, M. A. (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis, *J Bacteriol* 176, 2654-2662.
106. Zettler, J., and Mootz, H. D. (2010) Biochemical evidence for conformational changes in the cross-talk between adenylation and peptidyl-carrier protein domains of nonribosomal peptide synthetases, *FEBS J* 277, 1159-1171.
107. Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M., and Escalante-Semerena, J. C. (2003) The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A, *Biochemistry* 42, 2866-2873.
108. Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A., and Stubbs, M. T. (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains, *J Biol Chem* 283, 32484-32491.
109. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases, *Structure* 8, 407-418.
110. Lai, J. R., Koglin, A., and Walsh, C. T. (2006) Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis, *Biochemistry* 45, 14869-14879.
111. Koglin, A., Mofid, M. R., Lohr, F., Schafer, B., Rogov, V. V., Blum, M. M., Mittag, T., Marahiel, M. A., Bernhard, F., and Dotsch, V. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases, *Science* 312, 273-276.
112. Koglin, A., Lohr, F., Bernhard, F., Rogov, V. V., Frueh, D. P., Strieter, E. R., Mofid, M. R.,

- Guntert, P., Wagner, G., Walsh, C. T., Marahiel, M. A., and Dotsch, V. (2008) Structural basis for the selectivity of the external thioesterase of the surfactin synthetase, *Nature* 454, 907-U968.
113. Miller, D. A., and Walsh, C. T. (2001) Yersiniabactin synthetase: probing the recognition of carrier protein domains by the catalytic heterocyclization domains, Cy1 and Cy2, in the chain-initiating HWMP2 subunit, *Biochemistry* 40, 5313-5321.
114. Balibar, C. J., Vaillancourt, F. H., and Walsh, C. T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains, *Chem Biol* 12, 1189-1200.
115. Samel, S. A. (2009) Arbeiten zur strukturellen Charakterisierung von Thioesterase-, Kondensations und Epimerisierungsdomänen nichtribosomaler Peptidsynthetasen, pp 1-146, Philipps-Universität, Marburg.
116. Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, *BMC Evol Biol* 7, 78.
117. Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nat Struct Biol* 9, 522-526.
118. Leslie, A. G. (1990) Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 Å resolution, *J Mol Biol* 213, 167-186.
119. Lewendon, A., Murray, I. A., Shaw, W. V., Gibbs, M. R., and Leslie, A. G. (1994) Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: evidence for a general base role for glutamate, *Biochemistry* 33, 1944-1950.
120. Stein, D. B., Linne, U., Hahn, M., and Marahiel, M. A. (2006) Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis, *Chembiochem* 7, 1807-1814.
121. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L. O. (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase, *Structure* 15, 781-792.
122. Du, L., and Lou, L. (2010) PKS and NRPS release mechanisms, *Nat Prod Rep* 27, 255-278.
123. Nardini, M., and Dijkstra, B. W. (1999) Alpha/beta hydrolase fold enzymes: the family keeps growing, *Curr Opin Struct Biol* 9, 732-737.
124. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., and Stubbs, M. T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE, *Structure* 10, 301-310.
125. Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T., and Wagner, G. (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase, *Nature* 454, 903-U962.
126. Akey, D. L., Kittendorf, J. D., Giraldez, J. W., Fecik, R. A., Sherman, D. H., and Smith, J. L.

- (2006) Structural basis for macrolactonization by the pikromycin thioesterase, *Nat Chem Biol* 2, 537-542.
127. Hoyer, K. M., Mahlert, C., and Marahiel, M. A. (2007) The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization, *Chem Biol* 14, 13-22.
  128. Duerfahrt, T., Eppelmann, K., Muller, R., and Marahiel, M. A. (2004) Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis, *Chem Biol* 11, 261-271.
  129. Gaitatzis, N., Kunze, B., and Muller, R. (2001) In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases, *Proc Natl Acad Sci U S A* 98, 11136-11141.
  130. Lesk, A. M. (1995) NAD-binding domains of dehydrogenases, *Curr Opin Struct Biol* 5, 775-783.
  131. Halo, L. M., Marshall, J. W., Yakasai, A. A., Song, Z., Butts, C. P., Crump, M. P., Heneghan, M., Bailey, A. M., Simpson, T. J., Lazarus, C. M., and Cox, R. J. (2008) Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires coexpression with an enoyl reductase, *Chembiochem* 9, 585-594.
  132. Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000) Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides, *Chem Biol* 7, 287-297.
  133. Hahn, M., and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains, *Proc Natl Acad Sci U S A* 101, 15585-15590.
  134. Linne, U., Doekel, S., and Marahiel, M. A. (2001) Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases, *Biochemistry* 40, 15824-15834.
  135. Strieker, M., Essen, L. O., Walsh, C. T., and Marahiel, M. A. (2008) Non-heme hydroxylase engineering for simple enzymatic synthesis of L-threo-hydroxyaspartic acid, *Chembiochem* 9, 374-376.
  136. Hausinger, R. P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes, *Crit Rev Biochem Mol Biol* 39, 21-68.
  137. Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K., and Schofield, C. J. (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase, *Nat Struct Biol* 7, 127-133.
  138. Meier, J. L., and Burkart, M. D. (2011) Proteomic analysis of polyketide and nonribosomal peptide biosynthesis, *Curr Opin Chem Biol* 15, 48-56.
  139. Evans, S. E., Williams, C., Arthur, C. J., Burston, S. G., Simpson, T. J., Crosby, J., and Crump, M. P. (2008) An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein,



*Chembiochem* 9, 2424-2432.

140. Roujeinikova, A., Simon, W. J., Gilroy, J., Rice, D. W., Rafferty, J. B., and Slabas, A. R. (2007) Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates, *J Mol Biol* 365, 135-145.
141. Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator, *Nucleic Acids Res* 35, W43-46.
142. Evan, G. I., Lewis, G. K., Ramsay, G., and Bishop, J. M. (1985) Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product, *Mol Cell Biol* 5, 3610-3616.
143. Gilje, B., Heikkila, R., Olteidal, S., Tjensvoll, K., and Nordgard, O. (2008) High-fidelity DNA polymerase enhances the sensitivity of a peptide nucleic acid clamp PCR assay for K-ras mutations, *J Mol Diagn* 10, 325-331.
144. Lannenpaa, M., Janonen, I., Holttä-Vuori, M., Gardemeister, M., Porali, I., and Sopanen, T. (2004) A new SBP-box gene BpSPL1 in silver birch (*Betula pendula*), *Physiol Plant* 120, 491-500.
145. Birnboim, H. C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Res* 7, 1513-1523.
146. Aljanabi, S. M., and Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques, *Nucleic Acids Res* 25, 4692-4693.
147. Remaut, E., Tsao, H., and Fiers, W. (1983) Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication, *Gene* 22, 103-113.
148. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol, *Nucleic Acids Res* 32, e115.
149. Roberts, R. J. (1976) Restriction endonucleases, *CRC Crit Rev Biochem* 4, 123-164.
150. Longley, M. J., and Mosbaugh, D. W. (1991) Characterization of DNA metabolizing enzymes in situ following polyacrylamide gel electrophoresis, *Biochemistry* 30, 2655-2664.
151. Ronaghi, M., Uhlen, M., and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate, *Science* 281, 363, 365.
152. Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation, *Nat Rev Microbiol* 2, 241-249.
153. Studier, F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system, *J Mol Biol* 219, 37-44.
154. Linne, U. (2001) Epimerisierungsdomänen aus Peptidsynthetasen - Untersuchungen zur Substratspezifität und zur Interaktion mit anderen Domänen, in *Fachbereich Chemie*, p 137, Philipps-Universität Marburg.
155. Hahn, M., and Stachelhaus, T. (2006) Harnessing the potential of communication-

mediating domains for the biocombinatorial synthesis of nonribosomal peptides, *Proceedings of the National Academy of Sciences of the United States of America* 103, 275-280.

156. Stachelhaus, T. (1994) Multifunktionelle Peptidsynthetasen: Modulstruktur und Domänen austausch, in *Fachbereich Chemie*, p 140, Philipps-Universität Marburg.
157. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation, *Nature* 258, 598-599.
158. Ueda, E. K., Gout, P. W., and Morganti, L. (2003) Current and prospective applications of metal ion-protein binding, *J Chromatogr A* 988, 1-23.
159. Gaberc-Porekar, V., and Menart, V. (2001) Perspectives of immobilized-metal affinity chromatography, *J Biochem Biophys Methods* 49, 335-360.
160. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem* 72, 248-254.
161. Shapiro, A. L., Vinuela, E., and Maizel, J. V., Jr. (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels, *Biochem Biophys Res Commun* 28, 815-820.
162. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227, 680-685.
163. Bennett, J., and Scott, K. J. (1971) Quantitative staining of fraction I protein in polyacrylamide gels using Coomassie brilliant blue, *Anal Biochem* 43, 173-182.
164. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods Mol Biol* 112, 531-552.
165. Adrian, M., Dubochet, J., Lepault, J., and McDowell, A. W. (1984) Cryo-electron microscopy of viruses, *Nature* 308, 32-36.
166. Karas, M., Bachmann, D., Bahr, U. and Hillenkamp, F. (1987) Matrix-assisted ultraviolet laser desorption of non-volatile compounds, *International Journal of Mass Spectrometry and Ion Processes* 78, 53-68.
167. Horvath, C. G., Preiss, B. A., and Lipsky, S. R. (1967) Fast liquid chromatography: an investigation of operating parameters and the separation of nucleotides on pellicular ion exchangers, *Anal Chem* 39, 1422-1428.
168. Dempster, A. J. (1918) A new Method of Positive Ray Analysis, *Physical Review* 11, 316-325.
169. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science* 246, 64-71.
170. Tomer, K. B., Arthur Moseley, M., Deterding, L. J., and Parker, C. E. (1994) Capillary liquid chromatography/mass spectrometry, *Mass Spectrometry Reviews* 13, 431-457.

171. Bogdanov, B., and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up, *Mass Spectrom Rev* 24, 168-200.
172. Ball, H. L., and Mascagni, P. (1996) Chemical synthesis and purification of proteins: a methodology, *Int J Pept Protein Res* 48, 31-47.
173. Liu, Y., and Bruner, S. D. (2007) Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases, *Chembiochem* 8, 617-621.
174. Redecke, L., von Bergen, M., Clos, J., Konarev, P. V., Svergun, D. I., Fittschen, U. E., Broekaert, J. A., Bruns, O., Georgieva, D., Mandelkow, E., Genov, N., and Betzel, C. (2007) Structural characterization of beta-sheeted oligomers formed on the pathway of oxidative prion protein aggregation in vitro, *J Struct Biol* 157, 308-320.
175. Frisken, B. J. (2001) Revisiting the method of cumulants for the analysis of dynamic light-scattering data, *Appl Opt* 40, 4087-4091.
176. Bernado, P., and Blackledge, M. (2010) Structural biology: Proteins in dynamic equilibrium, *Nature* 468, 1046-1048.
177. Lipfert, J., and Doniach, S. (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes, *Annu Rev Biophys Biomol Struct* 36, 307-327.
178. Forster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A., and Sali, A. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies, *J Mol Biol* 382, 1089-1106.
179. Koch, M. H., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution, *Q Rev Biophys* 36, 147-227.
180. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, *Q Rev Biophys* 40, 191-285.
181. Guinier, A. a. F., G. (1955) Small angle scattering of X-rays., Chapter 4.
182. Porod, G. (1951) X-ray low angle scattering of dense colloid systems, *Kolloid Zeitschrift*, 83-114.
183. Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C., and Svergun, D. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg, *Journal of Applied Crystallography* 40, s190-s194.
184. Round, A. R., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D. I., and Roessle, M. (2008) Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33, *Journal of Applied Crystallography* 41, 913-917.
185. Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briesse, C., Suzuki, M., Tomizaki, T., Toyokawa, H., and Wagner, A. (2006) The PILATUS 1M detector, *J Synchrotron Radiat* 13, 120-130.

186. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis, *Journal of Applied Crystallography* 39, 277-286.
187. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis, *Journal of Applied Crystallography* 40, s223-s228.
188. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *Journal of Applied Crystallography* 25, 495-503.
189. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *Journal of Applied Crystallography* 36, 1277-1282.
190. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *Journal of Applied Crystallography* 42, 342-346.
191. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *Journal of Applied Crystallography* 28, 768-773.
192. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophys J* 89, 1237-1250.
193. von Ardenne, M. (1938) Das Elektronen-Rastermikroskop, *Zeitschrift für Physik A Hadrons and Nuclei* 109, 553-572.
194. Baumeister, W., and Steven, A. C. (2000) Macromolecular electron microscopy in the era of structural genomics, *Trends Biochem Sci* 25, 624-631.
195. Sander, B., Golas, M. M., Luhrmann, R., and Stark, H. (2010) An approach for de novo structure determination of dynamic molecular assemblies by electron cryomicroscopy, *Structure* 18, 667-676.
196. Levitt, M. (2007) Growth of novel protein structural data, *Proc Natl Acad Sci U S A* 104, 3183-3188.
197. Reis, N. M., Chirgadze, D. Y., Blundell, T. L., and Mackley, M. R. (2009) The effect of protein-precipitant interfaces and applied shear on the nucleation and growth of lysozyme crystals, *Acta Crystallographica Section D* 65, 1127-1139.
198. Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures, *FEBS J* 275, 1-21.
199. Budayova-Spano, M., Dauvergne, F., Audiffren, M., Bactivelane, T., and Cusack, S. (2007) A methodology and an instrument for the temperature-controlled optimization of crystal growth, *Acta Crystallographica Section D* 63, 339-347.
200. Kiefersauer, R., Than, M. E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J. M.,

- Soulimane, T., and Huber, R. (2000) A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes, *Journal of Applied Crystallography* 33, 1223-1230.
201. Alcorn, T., and Juers, D. H. (2010) Progress in rational methods of cryoprotection in macromolecular crystallography, *Acta Crystallographica Section D* 66, 366-373.
202. Watanabe, N. (2006) From phasing to structure refinement in-house: Cr/Cu dual-wavelength system and a loopless free crystal-mounting method, *Acta Crystallographica Section D* 62, 891-896.
203. Kabsch, W. (2010) XDS, *Acta Crystallographica Section D* 66, 125-132.
204. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat Struct Biol* 4, 269-275.
205. Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr* 62, 72-82.
206. McCoy, A. J., and Read, R. J. (2010) Experimental phasing: best practice and pitfalls, *Acta Crystallographica Section D* 66, 458-469.
207. Berman, H. (2008) The Protein Data Bank: a historical perspective, *Acta Crystallographica Section A* 64, 88-95.
208. Taylor, G. (2010) Introduction to phasing, *Acta Crystallographica Section D* 66, 325-338.
209. Joyce, M. G., Radaev, S., and Sun, P. D. (2010) A rational approach to heavy-atom derivative screening, *Acta Crystallographica Section D* 66, 358-365.
210. Skubak, P., Murshudov, G., and Pannu, N. S. (2009) A multivariate likelihood SIRAS function for phasing and model refinement, *Acta Crystallographica Section D* 65, 1051-1061.
211. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP, *Acta Crystallographica Section D* 66, 22-25.
212. McCoy, A. (2007) Solving structures of protein complexes by molecular replacement with Phaser, *Acta Crystallographica Section D* 63, 32-41.
213. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography* 40, 658-674.
214. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J Mol Biol* 302, 205-217.
215. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res* 22, 4673-4680.
216. Poirot, O., Suhre, K., Abergel, C., O'Toole, E., and Notredame, C. (2004) 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence

alignment, *Nucleic Acids Res* 32, W37-40.

217. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006) Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee, *Nucleic Acids Res* 34, W604-608.
218. Snook, C. F., Purdy, M. D., and Wiener, M. C. (2000) Use of a crystallization robot to set up sitting-drop vapor-diffusion crystallization and in situ crystallization screens, *Journal of Applied Crystallography* 33, 344-349.
219. Tanovic, A. (2007) Rationale Mutagenese nichtribosomaler Peptidsynthetasen und erste strukturelle Untersuchungen konformationserstarrter Module, p 112, Philipps-Universität, Marburg.
220. Cowtan, K. D., and Main, P. (1996) Phase combination and cross validation in iterated density-modification calculations, *Acta Crystallographica Section D* 52, 43-48.
221. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica Section D* 53, 240-255.
222. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D* 66, 486-501.
223. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography* 26, 283-291.
224. Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R., and Nureki, O. (2011) Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog, *Proc Natl Acad Sci U S A* 108, 3912-3917.
225. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison, *Bioinformatics* 16, 566-567.
226. Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J., and Edelman, M. (2005) SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment, *Nucleic Acids Res* 33, W39-43.
227. Koglin, A., Mofid, M. R., Lohr, F., Schafer, B., Rogov, V. V., Blum, M. M., Mittag, T., Marahiel, M. A., Bernhard, F., and Dotsch, V. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases - Supporting Online Material, *Science* 312, 273-276.
228. Weissman, K. J., and Muller, R. (2008) Protein-protein interactions in multienzyme megasynthetases, *ChemBiochem* 9, 826-848.
229. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A* 98, 10037-10041.
230. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
231. Bergendahl, V., Linne, U., and Marahiel, M. A. (2002) Mutational analysis of the C-



- domain in nonribosomal peptide synthesis, *Eur J Biochem* 269, 620-629.
232. Young, W. S., and Brooks, C. L., 3rd. (1996) A microscopic view of helix propagation: N and C-terminal helix growth in alanine helices, *J Mol Biol* 259, 560-572.
233. Couch, V. A., Cheng, N., Nambiar, K., and Fink, W. (2006) Structural characterization of alpha-helices of implicitly solvated poly-alanine, *J Phys Chem B* 110, 3410-3419.
234. Fierz, B., Reiner, A., and Kiefhaber, T. (2009) Local conformational dynamics in alpha-helices measured by fast triplet transfer, *Proc Natl Acad Sci U S A* 106, 1057-1062.
235. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., and Lundegaard, C. (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions, *BMC Struct Biol* 9, 51.
236. Tanovic, A., Samel, S. A., Essen, L. O., and Marahiel, M. A. (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase, *Science* 321, 659-663.
237. Leibundgut, M., Jenni, S., Frick, C., and Ban, N. (2007) Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase, *Science* 316, 288-290.
238. Arnold, K., Kiefer, F., Kopp, J., Battey, J. N., Podvinec, M., Westbrook, J. D., Berman, H. M., Bordoli, L., and Schwede, T. (2009) The Protein Model Portal, *J Struct Funct Genomics* 10, 1-8.
239. Amoutzias, G. D., Van De Peer, Y., and Mossialos, D. (2008) Evolution and taxonomic distribution of nonribosomal peptide and polyketide synthases, *Future Microbiology* 3, 361-370.
240. Baltz, R. H., Miao, V., and Wrigley, S. K. (2005) Natural products to drugs: daptomycin and related lipopeptide antibiotics, *Nat Prod Rep* 22, 717-741.
241. Rouhiainen, L., Jokela, J., Fewer, D. P., Urmann, M., and Sivonen, K. (2010) Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria), *Chem Biol* 17, 265-273.
242. Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M., and Kanaya, S. (2003) Cloning and characterization of the gene cluster encoding arthrofactin synthetase from *Pseudomonas* sp. MIS38, *Chem Biol* 10, 869-880.
243. Slightom, J. L., Metzger, B. P., Luu, H. T., and Elhammer, A. P. (2009) Cloning and molecular characterization of the gene encoding the Aureobasidin A biosynthesis complex in *Aureobasidium pullulans* BP-1938, *Gene* 431, 67-79.
244. May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001) The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *J Biol Chem* 276, 7209-7217.
245. Wu, X., Ballard, J., and Jiang, Y. W. (2005) Structure and biosynthesis of the BT peptide antibiotic from *Brevibacillus texasporus*, *Appl Environ Microbiol* 71, 8519-8530.
246. Magarvey, N. A., Ehling-Schulz, M., and Walsh, C. T. (2006) Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: Activation of alpha-keto acids

and chiral reduction on the assembly line, *Journal of the American Chemical Society* 128, 10698-10699.

247. von Dohren, H., Keller, U., Vater, J., and Zocher, R. (1997) Multifunctional Peptide Synthetases, *Chem Rev* 97, 2675-2706.
248. Velkov, T., and Lawen, A. (2003) Non-ribosomal peptide synthetases as technological platforms for the synthesis of highly modified peptide bioeffectors--Cyclosporin synthetase as a complex example, *Biotechnol Annu Rev* 9, 151-197.
249. Yin, X. H., and Zabriskie, T. M. (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*, *Microbiology-Sgm* 152, 2969-2983.
250. Roche, E. D., and Walsh, C. T. (2003) Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis, *Biochemistry* 42, 1334-1344.
251. Robbel, L., Knappe, T. A., Linne, U., Xie, X., and Marahiel, M. A. (2010) Erythrochelin--a hydroxamate-type siderophore predicted from the genome of *Saccharopolyspora erythraea*, *FEBS J* 277, 663-676.
252. Binz, T. M., Maffioli, S. I., Sosio, M., Donadio, S., and Muller, R. (2010) Insights into an unusual nonribosomal peptide synthetase biosynthesis: identification and characterization of the GE81112 biosynthetic gene cluster, *J Biol Chem* 285, 32710-32719.
253. Wang, Y., Chen, Y., Shen, Q., and Yin, X. (2011) Molecular cloning and identification of the laspartomycin biosynthetic gene cluster from *Streptomyces viridochromogenes*, *Gene* 483, 11-21.
254. Yakimov, M. M., Kroger, A., Slepak, T. N., Giuliano, L., Timmis, K. N., and Golyshin, P. N. (1998) A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization, *Biochim Biophys Acta* 1399, 141-153.
255. Krishnamurthy, T., Szafraniec, L., Hunt, D. F., Shabanowitz, J., Yates, J. R., 3rd, Hauer, C. R., Carmichael, W. W., Skulberg, O., Codd, G. A., and Missler, S. (1989) Structural characterization of toxic cyclic peptides from blue-green algae by tandem mass spectrometry, *Proc Natl Acad Sci U S A* 86, 770-774.
256. Nishizawa, T., Ueda, A., Nakano, T., Nishizawa, A., Miura, T., Asayama, M., Fujii, K., Harada, K., and Shirai, M. (2011) Characterization of the locus of genes encoding enzymes producing heptadepsipeptide micropeptin in the unicellular cyanobacterium *Microcystis*, *J Biochem* 149, 475-485.
257. Gunsior, M., Breazeale, S. D., Lind, A. J., Ravel, J., Janc, J. W., and Townsend, C. A. (2004) The biosynthetic gene cluster for a monocyclic beta-lactam antibiotic, nocardicin A, *Chem Biol* 11, 927-938.
258. Clardy, J., Fischbach, M. A., and Walsh, C. T. (2006) New antibiotics from bacterial natural products, *Nat Biotechnol* 24, 1541-1550.
259. de Crecy-Lagard, V., Saurin, W., Thibaut, D., Gil, P., Naudin, L., Crouzet, J., and Blanc, V. (1997) Streptogramin B biosynthesis in *Streptomyces pristinaespiralis* and

- Streptomyces virginiae*: molecular characterization of the last structural peptide synthetase gene, *Antimicrob Agents Chemother* 41, 1904-1909.
260. Dubern, J. F., Coppoolse, E. R., Stiekema, W. J., and Bloemberg, G. V. (2008) Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in *Pseudomonas putida* strain PCL1445, *Microbiology* 154, 2070-2083.
261. Qu, X., Jiang, N., Xu, F., Shao, L., Tang, G., Wilkinson, B., and Liu, W. (2011) Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor, *Mol Biosyst* 7, 852-861.
262. Scholz-Schroeder, B. K., Soule, J. D., and Gross, D. C. (2003) The *sypA*, *sypS*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D, *Mol Plant Microbe Interact* 16, 271-280.
263. Cheng, Y. Q. (2006) Deciphering the biosynthetic codes for the potent anti-SARS-CoV cyclodepsipeptide valinomycin in *Streptomyces tsusimaensis* ATCC 15141, *Chembiochem* 7, 471-477.
264. Christiansen, G., Philmus, B., Hemscheidt, T., and Kurmayer, R. (2011) Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity, *J Bacteriol* 193, 3822-3831.
265. Thomas, M. G., Chan, Y. A., and Ozanick, S. G. (2003) Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster, *Antimicrob Agents Chemother* 47, 2823-2830.
266. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, *Curr Protoc Bioinformatics Chapter 5*, Unit 5 6.
267. Eswar, N., Eramian, D., Webb, B., Shen, M. Y., and Sali, A. (2008) Protein structure modeling with MODELLER, *Methods Mol Biol* 426, 145-159.
268. Bond, C. S., and Schuttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments, *Acta Crystallogr D Biol Crystallogr* 65, 510-512.
269. Perham, R. N., Duckworth, H. W., and Roberts, G. C. (1981) Mobility of polypeptide chain in the pyruvate dehydrogenase complex revealed by proton NMR, *Nature* 292, 474-477.
270. Roberts, E. L., Shu, N., Howard, M. J., Broadhurst, R. W., Chapman-Smith, A., Wallace, J. C., Morris, T., Cronan, J. E., Jr., and Perham, R. N. (1999) Solution structures of apo and holo biotinyl domains from acetyl coenzyme A carboxylase of *Escherichia coli* determined by triple-resonance nuclear magnetic resonance spectroscopy, *Biochemistry* 38, 5045-5053.
271. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases, *Science* 284, 482-485.

272. Jenni, S., Leibundgut, M., Maier, T., and Ban, N. (2006) Architecture of a fungal fatty acid synthase at 5 Å resolution, *Science* **311**, 1263-1267.
273. Liu, Y., Zheng, T., and Bruner, S. D. (2011) Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases, *Chem Biol* **18**, 1482-1488.
274. Damborsky, J., Petrek, M., Banas, P., and Otyepka, M. (2007) Identification of tunnels in proteins, nucleic acids, inorganic materials and molecular ensembles, *Biotechnol J* **2**, 62-67.
275. Wattana-amorn, P., Williams, C., Ploskon, E., Cox, R. J., Simpson, T. J., Crosby, J., and Crump, M. P. (2010) Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase, *Biochemistry* **49**, 2186-2193.
276. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Res* **32**, W665-667.
277. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic Acids Res* **35**, W522-525.
278. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *Journal of the American Chemical Society* **117**, 5179-5197.
279. Hendle, J., Mattevi, A., Westphal, A. H., Spee, J., de Kok, A., Teplyakov, A., and Hol, W. G. (1995) Crystallographic and enzymatic investigations on the role of Ser558, His610, and Asn614 in the catalytic mechanism of *Azotobacter vinelandii* dihydrolipoamide acetyltransferase (E2p), *Biochemistry* **34**, 4287-4298.
280. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* **25**, 1605-1612.
281. Chiocchini, C., Linne, U., and Stachelhaus, T. (2006) In vivo biocombinatorial synthesis of lipopeptides by COM domain-mediated reprogramming of the surfactin biosynthetic complex, *Chemistry & Biology* **13**, 899-908.
282. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) Construction of hybrid peptide synthetases by module and domain fusions, *Proc Natl Acad Sci U S A* **97**, 5848-5853.
283. Richter, C. D., Nietlispach, D., Broadhurst, R. W., and Weissman, K. J. (2008) Multienzyme docking in hybrid megasynthetases, *Nature Chemical Biology* **4**, 75-81.
284. Luo, L., Burkart, M. D., Stachelhaus, T., and Walsh, C. T. (2001) Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase, *J Am Chem Soc* **123**, 11208-11218.
285. Adams, M. W., Dailey, H. A., DeLucas, L. J., Luo, M., Prestegard, J. H., Rose, J. P., and

- Wang, B. C. (2003) The Southeast Collaboratory for Structural Genomics: a high-throughput gene to structure factory, *Acc Chem Res* 36, 191-198.
286. Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L. S., Miller, M. D., McPhillips, T. M., Miller, M. A., Scheibe, D., Canaves, J. M., Guda, C., Jaroszewski, L., Selby, T. L., Elsliger, M. A., Wooley, J., Taylor, S. S., Hodgson, K. O., Wilson, I. A., Schultz, P. G., and Stevens, R. C. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, *Proc Natl Acad Sci U S A* 99, 11664-11669.
287. Yu, Y., Smith, D. M., Kim, H. M., Rodriguez, V., Goldberg, A. L., and Cheng, Y. (2010) Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions, *EMBO J* 29, 692-702.
288. Bulkley, D., Innis, C. A., Blaha, G., and Steitz, T. A. (2010) Revisiting the structures of several antibiotics bound to the bacterial ribosome, *Proc Natl Acad Sci U S A* 107, 17158-17163.
289. Orru, R., Dudek, H. M., Martinoli, C., Torres Pazmino, D. E., Royant, A., Weik, M., Fraaije, M. W., and Mattevi, A. (2011) Snapshots of enzymatic Baeyer-Villiger catalysis: oxygen activation and intermediate stabilization, *J Biol Chem* 286, 29284-29291.
290. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization, *Structure* 14, 1617-1622.
291. Wenzel, T., and Baumeister, W. (1995) Conformational constraints in protein degradation by the 20S proteasome, *Nat Struct Biol* 2, 199-204.
292. Jahn, W. (1999) Review: chemical aspects of the use of gold clusters in structural biology, *J Struct Biol* 127, 106-112.
293. Segond von Banchet, G., and Heppelmann, B. (1995) Non-radioactive localization of substance P binding sites in rat brain and spinal cord using peptides labeled with 1.4-nm gold particles, *J Histochem Cytochem* 43, 821-827.
294. Samel, S. A., Wagner, B., Marahiel, M. A., and Essen, L. O. (2006) The thioesterase domain of the fengycin biosynthesis cluster: A structural base for the macrocyclization of a non-ribosomal lipopeptide, *Journal of Molecular Biology* 359, 876-889.
295. Claxton, H. B., Akey, D. L., Silver, M. K., Admiraal, S. J., and Smith, J. L. (2009) Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway, *J Biol Chem* 284, 5021-5029.



## DANKSAGUNG

---

Mein besonderer Dank gilt Herrn Prof. Dr. Mohamed A. Marahiel für die interessante Themenstellung, sein Interesse an meiner Arbeit und seine stete Unterstützung. Desweiteren möchte ich ihm für die Möglichkeit am Teilchenbeschleuniger in Hamburg (D), Grenoble (F) und Villigen (CH) zu Messen und mich mit anderen Kollegen auszutauschen, herzlich danken.

Bei Herrn Prof. Klaus Reuter bedanke ich mich herzlich für die Übernahme des Zweitgutachtens.

Herrn Prof. Dr. Lars-Oliver Essen danke ich für die erfolgreiche Zusammenarbeit sowie die Einblicke in die faszinierende Welt der Kristallstrukturaufklärung. Darüber hinaus möchte ich ihm, und auch seinem ganzem Arbeitskreis, für die zahlreichen, interessanten Gespräche und Anregungen danken.

Dr. Uwe Linne danke ich für seine stete Unterstützung sowie Hilfe bei massenspektrometrischen Fragestellungen und Messungen.

Dr. Michal Gajda und Dr. Dennis Thomas möchte ich für die Zusammenarbeit bei der Aufnahme und Auswertung der SAXS- bzw. CryoEM-Daten danken.

Für die technische und organisatorische Unterstützung möchte ich mich bei Antje Schäfer, Christiane Bomm, Gabriele Schimpff-Weiland, Roswitha Roller-Müller, Carola Kisselbach, Sebastian Friedrich und Anke Botthoff bedanken.

Meinen Vertiefungsstudenten Simon Muche und Sven DeCausmaeker, meiner Bachelorstudentin Elisabeth Martin sowie meinem Diplomstudenten Volker Gatterdam möchte ich mich für ihre engagierte Mitarbeit und hohe Lernbereitschaft danken.

Für den regen Austausch und Unterstützung möchte ich mich herzlich bei meinen Kollegen Verena Helmetag, Florian Pecukert, Matthias Strieker, Thomas Knappe und Lars Robbel bedanken. Mein Dank gilt zusätzlich meinen Laborkollegen Jie Hou und Ana Laura Ramos Vega, sowie Femke Kraas und Tobias Giessen für die Korrektur meiner Arbeit.

Für die fachübergreifende Unterstützung möchte ich mich herzlich bei meinen Freunden und Kommilitonen Robert Grzimek und Ulrich Welling bedanken.

Mein größter Dank gebührt meiner Super-Frau Kerstin.